OM protein - protein search, using sw model

Copyright (c) 1993 - 2000 Compugen Ltd.

GenCore version 4.5

Run on: October 18, 2001, 09:08:45 ; Search time 22.23 Seconds
(without alignments)
1219.025 Million cell updates/sec

Title: US-09-596-958-2
Perfect score: 2310
Sequence: 1 MSILTLNNTSSSPGLFQSG.........LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

/Database : A_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	2310	100.0	447	20	AAW96260	Hypersensitive res
2	2310	100.0	447	21	AAY71094	Erwinia amylovora
3	2310	100.0	447	21	AAY84855	A hypersensitive r
4	559	24.2	424	20	AAW97851	Hypersensitive res
5	559	24.2	424	21	AAY71098	Pseudomonas syring
6	559	24.2	424	21	AAY84859	A hypersensitive r
7	359.5	15.6	197	19	AAW83014	Bacillus sp strain
8	359.5	15.6	197	19	AAW77412	Bacillus sp. pecti
9	341	14.8	221	20	AAY28446	Bacillus lichenifo
10	181.5	7.9	2411	21	AAB23860	Haemophilus influe
11	179	7.7	730	21	AAG29582	Arabidopsis thalia
12	179	7.7	752	21	AAG29581	Arabidopsis thalia
13	174	7.5	2042	19	AAW56319	Haemophilus paraga
14	165.5	7.2	344	18	AAW06600	Hypersensitive res
15	165.5	7.2	344	19	AAW75865	Pseudomonas solana
16	165.5	7.2	344	19	AAW62457	Pseudomonas solana
17	165.5	7.2	344	19	AAW61116	Hypersensitive res
18	165.5	7.2	344	20	AAW87641	A hypersensitive r
19	165.5	7.2	344	21	AAY71099	Pseudomonas solana
20	165.5	7.1	344	21	AAY84860	A hypersensitive r
21	164.5	7.1	898	18	AAW31853	Mycobacterium tube
22	157.5	6.8	2870	21	AAY95559	Caenorhabditis ele
23	157.5	6.8	3178	21	AAY95556	Caenorhabditis ele
24	156	6.8	2039	19	AAW56322	Haemophilus paraga
25	155.5	6.7	2353	17	AAR99393	Haemophilus adhesi
26	155	6.7	385	15	AAR45751	Erwinia amylovora
27	155	6.7	495	19	AAW39747	Human marcoSR prot
28	151	6.5	385	18	AAW06598	Hypersensitive res
29	151	6.5	403	19	AAW75863	Erwinia amylovora
30	151	6.5	403	19	AAW62455	Erwinia amylovora
31	151	6.5	403	19	AAW61114	Hypersensitive res
32	151	6.5	403	20	AAW87639	A hypersensitive r
33	151	6.5	403	21	AAY71093	Erwinia amylovora
34	151	6.5	403	21	AAY84854	A hypersensitive r
35	151	6.5	889	19	AAW68208	M. catarrhalis str
36	150	6.5	520	19	AAW39748	Human marcoSR prot
37	150	6.5	520	21	AAB44332	Human PRO1559 prot
38	150	6.5	871	21	AAY95550	Chlamydia pneumoni
39	148.5	6.4	885	21	AAY90238	Mature Chlamydia a
40	148.5	6.4	928	21	AAY90237	Chlamydia antigen
41	148	6.4	440	21	AAY99354	Human PRO1411 (UNQ
42	148	6.4	440	22	AAB87551	Human PRO1411. Ho
43	148	6.4	440	22	AAB31205	Amino acid sequenc
44	148	6.4	440	22	AAB66103	Protein of the inv
45	147.5	6.4	403	21	AAB58351	Lung cancer associ

ALIGNMENTS

RESULT 1
AAW96260

ID   AAW96260 standard; Protein; 447 AA.
XX
AC   AAW96260;
XX
DT   14-JUN-1999 (first entry)
XX
DE   Hypersensitive response eliciting protein (HrpW).
XX
KW   Hypersensitive response elicitor protein; hypersensitive response;
KW   hrpW; pathogen; infection; crop protection; disease resistance;
KW   pest resistance; transgenic plant; colouration; maturation.
XX
OS   Erwinia amylovora.
XX
PN   WO9907208-A1.
XX
PD   18-FEB-1999.
XX
PF   27-JUL-1998; 98WO-US15547.
XX
PR   06-AUG-1997; 97US-0055108.
XX
PA   (CORR) CORNELL RES FOUND INC.
XX
PI   Beer SV, Kim JF;
XX
DR   WPI; 1999-167126/14.
DR   N-PSDB; AAX09007.
XX
PT   New Erwinia amylovora hypersensitive response eliciting gene and
PT   protein - useful for providing transgenic plants and seeds with
PT   enhanced growth, and insect and disease resistance

```
XX
PS   Claim 1; Page 50-51; 54pp; English.
XX
CC   The hypersensitive response eliciting protein (hrp) or polypeptide
CC   is produced as part of an active defense by plants against
CC   incompatible pathogen infections. The hypersensitive response is a
CC   rapid localised necrosis. The hrp protein and gene when used in
CC   nucleotide constructs are useful for providing disease resistance to
CC   plants, insect control to plants, and enhancing plant growth
CC   (enhancing fruit size and earlier colouration and maturation), by
CC   direct application of the protein to plants, or by producing
CC   transgenic plants or seeds using the hrp gene.
XX
SQ   Sequence  447 AA;

Query Match              100.0%;   Score 2310; DB 20;  Length 447;
Best Local Similarity    100.0%;   Pred. No. 3.1e-152;
Matches  447;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MSILTLNNTSSSPGLFQSSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  msiltlnnntsssspglfqsggdnglgghnansalgqqpidrqtieqmaqllaellkslls   60

Qy   61  PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  pqsgnaatgaggndqttgvgnagglngrkgtagttpqsdsqnmlsemgnngldqaltpdg  120

Qy  121  QGGGQIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  qgggqigdnpllkamlkliarmdgqsdqfgqpgtgnnsassgtsssggspfndlsggka  180

Qy  181  PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  psgnspsgnyspvstfsppstptspldfpssptkaaggstpvtdhpdpvgsagigag  240

Qy  241  NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  nsvaftsaganqtvlhdtitvkagqvfdgkggtftagselgdggqsenqkplfiledgas  300

Qy  301  LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  lknvtmgddgadgihlygdakidnlhvtnvgedaitvkpnsagkkshveitnssfehasd  360

Qy  361  KILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLN  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  kilqlnadtnlsvdnvkakdfgtfvrtnggqqgnwdlnlshisaedgkfsfvksdsegln  420

Qy  421  VNTSDISLGDVENHYKVPMSANLKVAE  447
         ||||||||||||||||||||||||||||
Db  421  vntsdislgdvenhykvpmsanlkvae  447

RESULT  2
AAY71094
ID   AAY71094 standard; Protein; 447 AA.
XX
AC   AAY71094;
XX
DT   08-SEP-2000  (first entry)
XX
DE   Erwinia amylovora hypersensitive response elicitor #2.
XX
KW   Hypersensitive response elicitor; environmental stress resistance;
KW   plant.
XX
OS   Erwinia amylovora.
XX
PN   WO20002S055-A2.
XX
PD   18-MAY-2000.
XX
PF   04-NOV-1999;  99WO-US26039.
```

```
XX
PR   05-NOV-1998;  98US-0107243.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Schading RL;
XX
DR   WPI;  2000-376566/32.
DR   N-PSDB;  AAD00669.
XX
PT   Application of a hypersensitive response elicitor protein to plants to
PT   impart stress resistance -
XX
PS   Disclosure; Page 10-12; 84pp; English.
XX
CC   The patent discloses a method to impart stress resistance to plants by
CC   applying a hypersensitive response elicitor in a non-infectious form to
CC   a plant or seed. The present sequence is a hypersensitive
CC   response elicitor protein from Erwinia amylovora. The protein is
CC   heat stable, protease sensitive and suppressed by inhibitors of plant
CC   metabolism. It is used to impart stress resistance to plants.
XX
SQ   Sequence  447 AA;

Query Match              100.0%;   Score 2310; DB 21;  Length 447;
Best Local Similarity    100.0%;   Pred. No. 3.1e-152;
Matches  447;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1  MSILTLNNTSSSPGLFQSSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  msiltlnnntsssspglfqsggdnglgghnansalgqqpidrqtieqmaqllaellkslls   60

Qy   61  PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  pqsgnaatgaggndqttgvgnagglngrkgtagttpqsdsqnmlsemgnngldqaltpdg  120

Qy  121  QGGGQIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  qgggqigdnpllkamlkliarmdgqsdqfgqpgtgnnsassgtsssggspfndlsggka  180

Qy  181  PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG  240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  psgnspsgnyspvstfsppstptspldfpssptkaaggstpvtdhpdpvgsagigag  240

Qy  241  NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  nsvaftsaganqtvlhdtitvkagqvfdgkggtftagselgdggqsenqkplfiledgas  300

Qy  301  LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD  360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  lknvtmgddgadgihlygdakidnlhvtnvgedaitvkpnsagkkshveitnssfehasd  360

Qy  361  KILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLN  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  kilqlnadtnlsvdnvkakdfgtfvrtnggqqgnwdlnlshisaedgkfsfvksdsegln  420

Qy  421  VNTSDISLGDVENHYKVPMSANLKVAE  447
         ||||||||||||||||||||||||||||
Db  421  vntsdislgdvenhykvpmsanlkvae  447

RESULT  3
AAY84855
ID   AAY84855 standard; Protein; 447 AA.
XX
AC   AAY84855;
XX
DT   08-AUG-2000  (first entry)
XX
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; insect control; disease resistance;
```

```
Qy   361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQOGNWDLNLSHISAEDGKFSFVKSDSEGLN 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 kilqlnadtnlsvdnvkakdfgtfvrtnggqqgnwdlnlshisaedgkfsfvksdsegln 420

Qy   421 VNTSDISLGDVENHYKVPMSANLKVAE 447
         ||||||||||||||||||||||||||
Db   421 vntsdislgdvenhykvpmsanlkvae 447
```

RESULT   4
AAW97851
ID   AAW97851 standard; Protein; 424 AA.
XX
AC   AAW97851;
XX
DT   07-JUN-1999 (first entry)
XX
DE   Hypersensitive response elicitor HrpW.
XX
KW   Hypersensitive response elicitor; dspE gene; HrpW;
KW   transgenic plant; disease resistance; insect resistance.
XX
OS   Pseudomonas syringae pv. tomato.
XX
FH   Key   Location/Qualifiers
FT   Domain   1..186
FT   /note= "hypersensitive response elicitor-like
FT   domain"
FT   Region   119..186
FT   /note= "region of 6 imperfect glycine-rich repeats"
FT   Domain   187..424
FT   /note= "C-terminal domain"
XX
PN   WO907207-A1.
XX
PD   18-FEB-1999.
XX
PF   24-JUL-1998; 98WO-US15501.
XX
PR   06-AUG-1997; 97US-0055107.
XX
PA   (CORR) CORNELL RES FOUND INC.
XX
PI   Alfano JR, Charkowski A, Collmer A;
XX
DR   WPI; 1999-167125/14.
DR   N-PSDB; AAX24347.
XX
PT   New hypersensitive response eliciting (dspE) gene and protein -
PT   useful for providing transgenic plants and seeds with enhanced
PT   growth, and insect and disease resistance
XX
PS   Claim 18; Page 38-39; 56pp; English.
XX
CC   This the amino acid sequence of hypersensitive response element
CC   (HRE) HrpW of Pseudomonas syringae pv. tomato DC3000, as deduced
CC   from the nucleotide sequence of an isolated DNA molecule (see
CC   AAX24347). HrpW includes an HRE-like domain that is rich in Gln,
CC   Ser and Gly and which includes 6 imperfect glycine-rich repeats
CC   with many acidic and polar residues that align with similar
CC   repeats in the HrpZ proteins of P. syringae pv. syringae and P.
CC   syringae pv. tomato. The amino acid sequence of this region
CC   suggests alternating beta-sheets and turns that may form a barrel
CC   structure. The C-terminal domain of HrpW is similar to several
CC   fungal and bacterial Pel proteins. The HrpW protein or isolated
CC   DNA molecule can be used to impart disease resistance to plants, to
CC   enhance plant growth and/or to control insects on plants. This is
CC   achieved by applying the HRE protein in a non-infectious form to
CC   plants or plant seeds. Alternatively, transgenic plants or plant
CC   seeds transformed with DNA encoding the HRE can be provided.
XX
SQ   Sequence   424 AA;
```

---

```
KW   hypersensitive response elicitor; plant growth; vegetable; crop;
KW   ornamental plant.
XX
OS   Erwinia amylovora.
XX
PN   WO20002452-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;  99WO-US23181.
XX
PR   05-OCT-1998;  98US-0103050.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Fan H,  Niggemeyer JL;
XX
DR   WPI; 2000-303745/26.
DR   N-PSDB; AAA14939.
XX
PT   Hypersensitive response elicitor polypeptides useful for imparting
PT   enhanced growth, disease resistance and insect resistance to plants,
PT   especially vegetables and ornamental flowers -
XX
PS   Disclosure; Page 12-13; 100pp; English.
XX
CC   The present sequence represents a hypersensitive response elicitor
CC   polypeptide. The specification describes hypersensitive response
CC   elicitor polypeptide fragments, which do not elicit a hypersensitive
CC   response. Instead, the proteins impart disease resistance to plants,
CC   enhance plant growth, and/or control insects. The polypeptide
CC   fragments may be used to these properties to plants. The plants which
CC   may be treated in this way include vegetables, crops and ornamental
CC   plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
CC   peanut, corn, potato, sweet potato, bean, pea, chicory, lettuce,
CC   endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC   broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC   carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC   citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC   tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC.  petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ   Sequence   447 AA;
```

```
Query Match             100.0%;  Score 2310; DB 21;  Length 447;
Best Local Similarity   100.0%;  Pred. No. 3.1e-152;
Matches  447;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 MSILTLNNNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 msiltlnnntsssspglfqsggdnglgghnansalgqqpidrqtieqmaqllaellkslls 60

Qy    61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQDSQNMLSEMGNNGLDQAITPDG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 pqsgnaatgaggndqttgvgnaggingrkgtagttpqdsqnmlsemgnngldqaitpdg 120

Qy   121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 qgggqigdnpllkamlkliarmmdgqsdqfgqpgtgnnsassgtsssggspfndlsggka 180

Qy   181 PSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 psgnspsgnyspvstfsppstptsptspldfpsptptspldfpsspptkaaggstpvtdhpdpvgsagigag 240

Qy   241 NSVAFTSAGANQTVLHDTITVKAGQVFDCKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 nsvaftsaganqtvlhdtitvkaggvfdgkgqtftagselgdggqsenqkplfiledgas 300

Qy   301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 lknvtmgddgadgihlygdakidnlhvtnvgedaitvkpnsagkkshveitnssfehasd 360
```

```
Query Match             24.2%;     Score 559; DB 20;  Length 424;
Best Local Similarity   36.1%;     Pred. No. 4.8e-31;
Matches 147; Conservative  54; Mismatches 132;   Indels 74;   Gaps 13;

Qy  37 QPIDRQTIEQMAQLLAELLKSL---LSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAG 93
        :||:: :||::|:|:| |: :|                               |||
Db  72 kpndsqs--niaklisalimsllqmltnsnkkgdtnqeqpdsqapfqnngglg------- 122

Qy  94 TTPQSDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMMDGQSDQFGQP 153
        ||::|| | ||| ||
Db 123 -tpsads------gggg-----tpdatgggg-gdtp---------------satggg 151

Qy 154 GTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPS 213
        : ||:| |||  | || |
Db 152 ggdtptatgggggggggggtptatggg---sggtptatgggeggvtpqitpql------a 200

Qy 214 SPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQT 273
        :|: :| |: ||||||| :|:: |||:|
Db 201 nprtsg----------tgsvsdtags----teqagkinvvkdtikvgagevfdghgat 245

Qy 274 FTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYG----DAKIDNLHVTN 329
        ||| :| | | :||:||||| :||: | |||:
Db 246 ftadksmgngdgenqkpmfelaegatlknvnigenevdgihvkaknaqevtidnvhaqn 305

Qy 330 VGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNG 389
        ||| |||||| : || ||||
Db 306 vgedlitvkgeggaavtnlniknssakgaddkvvqlnanthlkidnfkaddfgtmvrtng 365

Qy 390 GQQ-GNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLGDVENHY 435
        :|: :|:|||| ||||| :|::| |:: :
Db 366 gkqfddmsielngieanhgkfalvksdsddlklatgniamtdvkhay 412


RESULT 5
AAY71098
ID   AAY71098 standard; Protein; 424 AA.
XX
AC   AAY71098;
XX
DT   08-SEP-2000  (first entry)
XX
DE   Pseudomonas syringae dspE gene encoded hypersensitive response elicitor.
XX
KW   Hypersensitive response elicitor; environmental stress resistance;
KW   plant; pathogen; dspE gene.
XX
OS   Pseudomonas syringae.
XX
PN   WO200028055-A2.
XX
PD   18-MAY-2000.
XX
PF   04-NOV-1999;   99WO-US26039.
XX
PR   05-NOV-1998;   98US-0107243.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,  Schading RL;
XX
DR   WPI; 2000-376566/32.
DR   N-PSDB; AAD00673.
XX
PT   Application of a hypersensitive response elicitor protein to plants to
PT   impart stress resistance -
XX
PS   Disclosure; Page 25-26; 84pp; English.
XX
CC   The patent discloses a method to impart stress resistance to plants by
CC   applying a hypersensitive response elicitor in a non-infectious form to
CC   a plant or seed. The present sequence is a hypersensitive response
CC   elicitor protein encoded by dspE gene from Pseudomonas syringae.
CC   The protein elicits a plant pathogen's hypersensitive response and is
CC   used to impart stress resistance to plants.
XX
SQ   Sequence  424 AA;


Query Match             24.2%;     Score 559; DB 21;  Length 424;
Best Local Similarity   36.1%;     Pred. No. 4.8e-31;
Matches 147; Conservative  54; Mismatches 132;   Indels 74;   Gaps 13;

Qy  37 QPIDRQTIEQMAQLLAELLKSL---LSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAG 93
        :||:: :||::|:|:| |: :|                               |||
Db  72 kpndsqs--niaklisalimsllqmltnsnkkgdtnqeqpdsqapfqnngglg------- 122

Qy  94 TTPQSDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMMDGQSDQFGQP 153
        ||::|| | ||| ||
Db 123 -tpsads------gggg-----tpdatgggg-gdtp---------------satggg 151

Qy 154 GTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPS 213
        : ||:| |||  | || |
Db 152 ggdtptatgggggggggggtptatggg---sggtptatgggeggvtpqitpql------a 200

Qy 214 SPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQT 273
        :|: :| |: ||||||| :|:: |||:|
Db 201 nprtsg----------tgsvsdtags----teqagkinvvkdtikvgagevfdghgat 245

Qy 274 FTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYG----DAKIDNLHVTN 329
        ||| :| | | :||:||||| :||: | |||:
Db 246 ftadksmgngdgenqkpmfelaegatlknvnigenevdgihvkaknaqevtidnvhaqn 305

Qy 330 VGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNG 389
        ||| |||||| : || ||||
Db 306 vgedlitvkgeggaavtnlniknssakgaddkvvqlnanthlkidnfkaddfgtmvrtng 365

Qy 390 GQQ-GNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLGDVENHY 435
        :|: :|:|||| ||||| :|::| |:: :
Db 366 gkqfddmsielngieanhgkfalvksdsddlklatgniamtdvkhay 412


RESULT 6
AAY84859
ID   AAY84859 standard; Protein; 424 AA.
XX
AC   AAY84859;
XX
DT   08-AUG-2000  (first entry)
XX
DE   A hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; insect control; disease resistance;
KW   hypersensitive response elicitor; plant growth; vegetable; crop;
KW   ornamental plant.
XX
OS   Pseudomonas syringae.
XX
PN   WO200020452-A2.
XX
PD   13-APR-2000.
XX
PF   05-OCT-1999;   99WO-US23181.
XX
PR   05-OCT-1998;   98US-0103050.
XX
PA   (EDEN-) EDEN BIOSCIENCE CORP.
XX
PI   Wei Z,   Fan H,   Niggemeyer JL;
XX
DR   WPI; 2000-303745/26.
DR   N-PSDB; AAA14943.
XX
PT   Hypersensitive response elicitor polypeptides useful for imparting
PT   enhanced growth, disease resistance and insect resistance to plants,
PT   especially vegetables and ornamental flowers -
```

```
XX
PS   Disclosure; Page 26-28; 100pp; English.
XX
CC   The present sequence represents a hypersensitive response elicitor
CC   polypeptide. The specification describes hypersensitive response
CC   elicitor polypeptide fragments, which do not elicit a hypersensitive
CC   response. Instead, the proteins impart disease resistance to plants,
CC   enhance plant growth, and/or control insects. The polypeptide
CC   fragments may be used to these properties to plants. The plants which
CC   may be treated in this way include vegetables, crops and ornamental
CC   plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
CC   peanut, corn, potato, sweet potato, bean, pea, chicory, lettuce,
CC   endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
CC   broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
CC   carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
CC   citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
CC   tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
CC   petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.
XX
SQ   Sequence   424 AA;

Query Match          24.2%;   Score 559;  DB 21;   Length 424;
Best Local Similarity  36.1%;  Pred. No. 4.8e-31;
Matches  147;  Conservative  54;  Mismatches 132;  Indels  74;  Gaps  13;

Qy   37 QPIDRQTIEQMAQLLAELLKSL---LSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAG 93
         :| :| | :|:||:| |: | |  | ::|     ||| |||||| | |||
Db   72 kpndsqs--niaklisalimsllqmltnsnkkgdtnqeqpdsqapfqnngglg------- 122

Qy   94 TTPQSDSQNMLSEMGNNGLDQAITPDGQGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQP 153
         || :||                ||| ||| |||
Db  123 -tpsads----------gggg-----tpdatggg-gdtp---------------satggg 151

Qy  154 GTGNNSASSCTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPS 213
         ||| || | ||||| ||| ||| |:|| | :|| ::  || :  ||| ::: | |
Db  152 ggdtptatgggsggggtptatggg--sggtptatgggeggvtpqitpql--------a 200

Qy  214 SPTKAAGGSTPVTDHPDPVGSAGTGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQT 273
         : |: :|    || : ||   |||  ||:|||| || || | | | ||:||||||||||
Db  201 npnrsg------------tgsvsdtags---teqagkinvvkdtikvqagevfdghgat 245

Qy  274 FTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYG---DAKIDNLHVTN 329
         ||| :||||||| | |||||||||||| | |||     |     || | ||:|||||
Db  246 ftadksmgngdqgqenqkpmfelaegatlknvlgenevdgihvkaknagevtidnvhaqn 305

Qy  330 VGEDAITVKPNSAGKKSHVEITNSSFEHASDKIIQLNADTNLSVDNVKAKDFGTFVRTNG 389
         |||| |||||||| ||||||| |:||:|  :||: | |   ||||||||  ||||| ||
Db  306 vgedlitvkgeggaavtniniknssakgaddkvvqlnanthlkidnfkaddfgtmvrtng 365

Qy  390 GQQ-GNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLGDVENHY 435
         |:| | ||||||| ::| : :|||||||  :| |  :: | |
Db  366 gkqfddmsielngleanhgkfalvksdsdiklatgniamtdvkhay 412

RESULT  7
AAW83014
ID   AAW83014 standard; Protein; 197 AA.
XX
AC   AAW83014;
XX
DT   28-JAN-1999  (first entry)
XX
DE   Bacillus sp strain KSM-P15 pectic acid lyase.
XX
KW   Bacillus sp. strain KSM-P15; pectic acid lyase; protopectinase;
KW   protopectin; polygalacturonic acid; detergent; cotton; surfactant;
KW   cellulase; protease; bleaching agent.
XX
OS   Bacillus sp.
XX
PN   WO9845393-A2.
XX
```

```
PD   15-OCT-1998.
XX
PF   08-APR-1998;  98WO-JP01613.
XX
PR   08-SEP-1997;  97JP-0242736.
PR   09-APR-1997;  97JP-0091142.
XX
PA   (KAOS ) KAO CORP.
XX
PI   Hatada Y, Ito S, Kasai M, Kobayashi T, Koike K;
PI   Shikata S, Suzumatsu A, Tsumadori M, Wada Y;
XX
DR   WPI; 1998-568339/48.
DR   N-PSDB; AAV69879.
XX
PT   Detergent composition containing protopectinase active at alkaline
PT   pH - on protopectin and polygalacturonic acid, provides better
PT   removal of muddy soil
XX
PS   Example; Page 71-72; 80pp; English.
XX
CC   The present invention describes a detergent composition which contains
CC   a protopectinase having an optimum pH 7 or higher against protopectin
CC   and polygalacturonic acid substrates. Inclusion of protopectinase
CC   gives a composition that provides better removal of muddy soil,
CC   particularly from socks. The present sequence represents pectic acid
CC   lyase from Bacillus sp. strain KSM-P15, which is used in an example
CC   from the present invention. Pectic acid lyase exhibits protopectinase
CC   activity.
XX
SQ   Sequence   197 AA;

Query Match          15.6%;   Score 359.5;  DB 19;   Length 197;
Best Local Similarity  46.3%;  Pred. No. 1.2e-17;
Matches   81;  Conservative  29;  Mismatches  56;  Indels   9;  Gaps   4;

Qy  253 TVLHDTITVKAGQVFDGKGQTFTAG-SELGDGGQSENQKPLFILEDGASLKNVTMGDDGA 311
         |:|:|| | || || |||||| :|  |:|||||||||||||||||||| :|  | :|  |
Db    3 tvvhetlrvpaggtfdgkgqtyvanptlgdsqaenqkplfrleagaslknvvigapaa 62

Qy  312 DGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNL 371
         :|   | |::| :|  |:|| | |:   :    |  | |  ||::| : | |
Db   63 dgvhcygdcitnviwedvgedaltlk--ssgt---vnisggaaykaydkvfqinaagti 117

Qy  372 SVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSE--GLNVNT 423
         ::: :| :| | || ||    :| :     ::|:|| |:|||||  | ||
Db  118 nirnfraddigklvrqnggtytykvvmvencisrvkdalirtdsststgrivnt 172

RESULT  8
AAW77412
ID   AAW77412 standard; Protein; 197 AA.
XX
AC   AAW77412;
XX
DT   07-JAN-1999  (first entry)
XX
DE   Bacillus sp. pectic acid lyase.
XX
KW   Bacillus sp. KSM-P15; pectic acid lyase; pectinic acid lyase; pectin;
KW   detergent; food-processing; fibre-processing agent; cell wall.
XX
OS   Bacillus sp.
XX
PN   EP870834-A1.
XX
PD   14-OCT-1998.
XX
PF   09-APR-1998;  98EP-0106586.
XX
PR   08-SEP-1997;  97JP-0242735.
PR   09-APR-1997;  97JP-0091142.
XX
```

```
XX
PA   (KAOS ) KAO CORP.
XX
PI   Hatada Y, Ito S, Kobayashi T, Koike K, Suzumatsu A;
PI   Yoshimatsu T;
XX
DR   WPI;.1998-523159/45.
DR   N-PSDB; AAV59478.
XX
PT   New Bacillus pectic acid lyase - useful as a detergent component, a
PT   food-processing agent and a fibre-processing agent
XX
PS   Claim 1; Page 16-17; 29pp; English.
XX
CC   The present sequence represents a pectic acid lyase isolated from
CC   microorganism Bacillus sp. KSM-P15. The pectic acid lyase has high
CC   pectic acid lyase activity which degrades pectin in plant cell walls
CC   and fibre in vegetables, and so is useful as a component of detergents,
CC   a food-processing agent, or a fibre-processing agent. The pectic acid
CC   lyase has a higher optimum reaction pH (10.3-10.7) than known Bacillus
CC   pectic acid lyases (pH 8-9.5) and so has wider industrial applications.
CC   Unlike present pectic acid lyases, the new enzyme has a high enzyme
CC   activity, and can be produced on a mass scale.
XX
SQ   Sequence   197 AA;

Query Match         15.6%;    Score 359.5; DB 19;   Length 197;
Best Local Similarity   46.3%;    Pred. No. 1.2e-17;
Matches   81;  Conservative  29; Mismatches  56;  Indels  9;  Gaps  4;

Qy   253 TVLHDTTTVKAGQVFDGKGQTFTAG-SELGDGGQSENQKPLFILEDGASLKNVTMGDDGA 311
         ||:|||| |||| |||||:| | : ||| ||||||:| |||||||||||||:|||||||
Db.    3 tvvhetirvpaggtfdgkggtyvanpntlgdgsqaenqkplfrleagaslknvvigapaa 62

Qy   312 DGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNL 371
         ||:||| || :|||||||||||| : |: : |: ||| : |||::| ||
Db    63 dgvhcygdctitnviwedvgedaltik--ssgt---vnisggaaykaydkvfqinaagti 117

Qy   372 SVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSE--GLNVNT 423
         : :|:|| |||||||||||||||||  ::||| :|||  |: || | |||
Db   118 nirnfraddigklvrqnggtgtykvvmvencnisrvkdailrtdsststgrivnt 172

RESULT 9
AAY28446
ID   AAY28446 standard; Protein; 221 AA.
XX
AC   AAY28446;
XX
DT   17-JAN-2000 (first entry)
XX
DE   Bacillus licheniformis Pectate lyase I.
XX
OS   Bacillus licheniformis.
XX
KW   Pectate lyase I; EC 4.2.2.2; pectin degrading enzyme; ATCC 14580;
KW   catalytic active domain; cellulose binding domain; CBD; operably linked;
KW   optimum activity; pH; detergent composition; yarn; cellulosic fibre;
KW   recycled waste paper; pulp; retting process; animal feed; wine; juice;
KW   transgenic plant.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..27
FT                   /label= Pro-sequence
FT   Protein         28..221
FT                   /label= Mature_pectate_lyase_I
FT   Misc-difference 133
FT                   /note= "Conserved residue"
FT   Misc-difference 155
FT                   /note= "Conserved residue"
XX
PN   WO9927083-A1.
```

```
XX
PD   03-JUN-1999.
XX
PF   98WO-DK00514.
XX
PR   24-NOV-1998;  98WO-DK00514.
PR   24-NOV-1997;  97DK-0001344.
PR   06-MAY-1998;  98US-0073684.
XX
PA   (NOVO ) NOVO-NORDISK AS.
XX
PI   Andersen LN, Bjornvad ME, Lange NEK, Schnorr K, Schuelein M;
XX
DR   WPI; 1999-610578/52.
DR   N-PSDB; AAX89484.
XX
PT   New isolated pectate lyase enzymes
XX
PS   Claim 6; Page 79-80; 93pp; English.
XX
CC   The present sequence is pectate lyase I, which is a pectin degrading
CC   enzyme derived from Bacillus licheniformis, ATCC 14580. The enzyme
CC   comprises a catalytic active domain and a cellulose binding domain
CC   (CBD) that are operably linked to each other. It shows optimum activity
CC   at pH greater than 9 and temperature 55 degree centigrade. It can be used
CC   in detergent compositions, for cleaning hard surfaces, for machine
CC   treatment of fabrics, for improving the properties of cellulosic fibres,
CC   yarn, woven or non-woven fabric, for the degradation of plant material
CC   e.g. recycled waste paper, mechanical paper-making pulps or fibres
CC   subjected to retting process, for preparing animal feed and for
CC   processing wine or juice. DNA encoding the enzyme can also be used for
CC   the production of transgenic plants.
XX
SQ   Sequence   221 AA;

Query Match          14.8%;    Score 341; DB 20;   Length 221;
Best Local Similarity  41.7%;   Pred. No. 2.6e-16;
Matches   73;  Conservative  33; Mismatches  59;  Indels  10;  Gaps  2;

Qy   254 VLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADG 313
         |:||| |||| |||||:|||||||||| |||||||:|||||||||||||:|||||||
Db    31 vvhktivvekgqtydgkgrllagpelgdgsqredqkplfkvedgatlknvvigapaadg 90

Qy   314 IHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSV 373
         :| |||:| :|||||||||||||  : :||| : |||::| ||| | :|
Db    91 vhtygnasinnvvwedvgedaltvk-----segsvtinggsarlaadklfqinkastftv 145

Qy   374 DNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISL 428
         | |: || |||||||||||||||  : :||| :|| : :|| |
Db   146 knftadqggkfirqlggstfkavvnidnctitnmkealfrtds----stssvtm 195

RESULT 10
AAB23860
ID   AAB23860 standard; Protein; 2411 AA.
XX
AC   AAB23860;
XX
DT   17-JAN-2001 (first entry)
XX
DE   Haemophilus influenzae adhesin (Hia) protein from type c strain API.
XX
KW   Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
KW   non-typeable Haemophilus influenzae; antiinflammatory; auditory;
KW   antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
KW   diagnosis; immunogenic; antigen.
XX
OS   Haemophilus influenzae.
XX
PN   WO200055191-A2.
XX
PD   21-SEP-2000.
XX
```

```
PF    16-MAR-2000; 2000WO-CA00289.                                    Qy   372 SVDNVKAKDFG------------------------------TFVRTNGGQQGN-- 394
XX                                                                              : |:|| |                                 | |:| ||
PR    16-MAR-1999; 99US-0268347.                                      Db  1615 tsvtveskdngkrtevkigaktsvikdhngklftgkelkdannngvtvtetdgkdegngl 1674
XX
PA    (CONN-) CONNAUGHT LAB LTD.                                      Qy   395 ------WDLNLSHISAEDGKFSFVKS------DSEGLNVNTSDISLGDVEN 433
XX                                                                                :          :: |||:|                | | :| ||
PI    Loosmore SM,  Yang Y,  Klein MH;                                Db  1675 vtakavidavnkagwrvkttgangqnddfatvasgtnvtfadgngttaevtkandgsitv 1734
XX
DR    WPI: 2000-618897/59.                                            Qy   434 HYKVPMSANLKV 445
DR    N-PSDB; AAA92499.                                                            | |: | ||
XX                                                                    Db  1735 kynvkvadglkl 1746
PT    Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT    use as antigens and vaccines and for treating Hemophilus influenzae
PT    infection -                                                     RESULT 11
XX                                                                    AAG29582
PS    Claim 1; Fig 24; 275pp; English.                               ID    AAG29582 standard; protein; 730 AA.
XX                                                                    XX
CC    The present sequence represents a Haemophilus influenzae adhesin AC    AAG29582;
CC    (Hia) protein from the type c Haemophilus influenzae strain API. XX
CC    Hia genes and proteins have antiinflammatory, auditory and antibacterial DT    17-OCT-2000  (first entry)
CC    activities, and can be used in the production of a vaccine. An   XX
CC    immunogenic composition comprising an Hia gene, a polypeptide encoded DE    Arabidopsis thaliana protein fragment SEQ ID NO: 35222.
CC    by an Hia gene, or a recombinant Hia polypeptide is useful for inducing XX
CC    protection against disease caused by Haemophilus strains in a    KW    Protein identification; signal transduction pathway; metabolic pathway;
CC    susceptible host, preferably a human. An Hia protein is useful as an KW    hybridisation assay; genetic mapping; gene expression control; promoter;
CC    antigen, in immunogenic preparations including vaccines, as a carrier KW    termination sequence.
CC    for other immunogens, and in the generation of diagnostic reagents. Hia XX
CC    is useful for treating diseases caused by the infection of Haemophilus OS    Arabidopsis thaliana.
CC    influenzae such as meningitis, epiglottitis, septicaemia and otitis XX
CC    media. Recombinant production of Hia favours high recovery of the PN    EP1033405-A2.
CC    protein compared to the low recovery of native protein from Haemophilus XX
CC    influenzae species. A truncated protein has a significantly higher PD    06-SEP-2000.
CC    amount of recovery than a full-length protein.                  XX
XX                                                                    PF    25-FEB-2000; 2000EP-0301439.
SQ    Sequence   2411 AA;                                             XX
                                                                      PR    25-FEB-1999; 99US-0121825.
                                                                      PR    05-MAR-1999; 99US-0123180.
Query Match         7.9%;  Score 181.5; DB 21;  Length 2411;         PR    09-MAR-1999; 99US-0123548.
Best Local Similarity  20.6%;  Pred. NO. 0.0006;                     PR    23-MAR-1999; 99US-0125788.
Matches  126; Conservative  64; Mismatches 205;  Indels 217;  Gaps  26; PR    25-MAR-1999; 99US-0126264.
                                                                      PR    29-MAR-1999; 99US-0126785.
Qy     4 LTLNN---NTSSSPGLFQSGGDNGLGGHN------------------ANSALGQQPIDRQT 43 PR    01-APR-1999; 99US-0127462.
          ||| |    :  |:: |||        ::                     |:||       PR    06-APR-1999; 99US-0128234.
Db  1182 itiangaaagtdasngnclsvtkdgisagnkeitnvksalktykdtqntagatcpa-ant 1240 PR    08-APR-1999; 99US-0128714.
                                                                      PR    16-APR-1999; 99US-0129845.
Qy    44 IEQMAQLLAELLKSLLSPOSGNNAATGAGGNDQTTGVGNAGGLNG-----RKGTAGTTPQ 97 PR    19-APR-1999; 99US-0130077.
          | ::| :||| :||| :|| :   |  ::|||  |            || |||        PR    21-APR-1999; 99US-0130449.
Db  1241 aevakqdlvdltk---patgaagngadakapdttaatvgdvgwvlsakktadetqd 1296 PR    23-APR-1999; 99US-0130510.
                                                                      PR    23-APR-1999; 99US-0130891.
Qy    98 SD------SQNMLSEMGNNGLDQAITPDGQG--------GGQIGDNPLLKAMLKLIARM 142 PR    28-APR-1999; 99US-0131449.
          ||      : ::||: || ||:| |             ||:|  :|              PR    30-APR-1999; 99US-0132048.
Db  1297 kefhaavknanevefvgkngatvsaktdnngkhtvtldvaeakvgdg--------lekd 1347 PR    30-APR-1999; 99US-0132407.
                                                                      PR    04-MAY-1999; 99US-0132484.
Qy   143 MDGQSDQFGQPGTGNN---SASSGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSP 198 PR    05-MAY-1999; 99US-0132485.
          |||            |||                  :: |||                  PR    06-MAY-1999; 99US-0132486.
Db  1348 tdgkiklkvdntdgnnlltvdatkgasva-------------------kgefnavtt--- 1385 PR    06-MAY-1999; 99US-0132487.
                                                                      PR    07-MAY-1999; 99US-0132863.
Qy   199 PSTPTSPTSPLDFPSSPTKAAGGSTPV-TDHPD--PVGSAGIGAGNSVAF---------- 245 PR    11-MAY-1999; 99US-0134256.
          : |:  ||||:  |  |||   |       ||                          PR    14-MAY-1999; 99US-0134218.
Db  1386 dattaqgtnanergkvvvkgsngatatetdkkkvatvgdvakaindaatfvkvenddsat 1445 PR    14-MAY-1999; 99US-0134219.
                                                                      PR    14-MAY-1999; 99US-0134221.
Qy   246 -----TSAGANQTV-LHDTITVKAGQVF---DGKGQTFTAGSE------------LGD 282 PR    14-MAY-1999; 99US-0134370.
          |||      |:::  |||:         |               |             PR    18-MAY-1999; 99US-0134768.
Db  1446 lddsptddgandalkagdtltlkagknlkvkrdgknitfalandlsvksatvsdklslgt 1505 PR    19-MAY-1999; 99US-0134941.
                                                                      PR    20-MAY-1999; 99US-0135124.
Qy   283 GGQSENQKPLFILEDGASL---KNVTMGDDGADGIHLYG-----DAKIDNLHVTNVGED 333 PR    21-MAY-1999; 99US-0135353.
          :|    ::| || |      |     |||     |:|  ||:|                PR    24-MAY-1999; 99US-0135629.
Db  1506 ngnkvn------itsdtkglnfaksktgdda-nlhlngiastltdtlinsgattnlggn 1558 PR    25-MAY-1999; 99US-0136021.
                                                                      PR    27-MAY-1999; 99US-0136392.
Qy   334 AIT-------------------------------VKPNSAGKKSHVEITNSSFEHASDKILQLNADTNL 371 PR    28-MAY-1999; 99US-0136782.
          ||                             ||:||  :: ||:|:|            PR    01-JUN-1999; 99US-0137222.
Db  1559 gltdnekkraasvkdvlnagwnvrgvkpasann---qvenidfvatydtvdfvsgdkdt 1614 PR    03-JUN-1999; 99US-0137528.
                                                                      PR    04-JUN-1999; 99US-0137502.
```

PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	99US-0139119.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139452.	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	99US-0139453.	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	25-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	26-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	27-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	30-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	3-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	10-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	13-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	15-SEP-1999;	99US-0153758.
PR	30-JUN-1999;	99US-0141287.	PR	16-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	20-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	22-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	23-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	24-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0155659.
PR	09-JUL-1999;	99US-0142920.	PR	29-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	04-OCT-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	05-OCT-1999;	99US-0157117.
PR	14-JUL-1999;	99US-0143624.	PR	06-OCT-1999;	99US-0157753.
PR	15-JUL-1999;	99US-0144005.	PR	07-OCT-1999;	99US-0157865.
PR	16-JUL-1999;	99US-0144085.	PR	08-OCT-1999;	99US-0158029.
PR	16-JUL-1999;	99US-0144086.	PR	12-OCT-1999;	99US-0158232.
PR	19-JUL-1999;	99US-0144325.	PR	13-OCT-1999;	99US-0158369.
PR	19-JUL-1999;	99US-0144331.	PR	13-OCT-1999;	99US-0159293.
PR	19-JUL-1999;	99US-0144332.	PR	14-OCT-1999;	99US-0159294.
PR	19-JUL-1999;	99US-0144333.	PR	14-OCT-1999;	99US-0159295.
PR	19-JUL-1999;	99US-0144334.	PR	14-OCT-1999;	99US-0159329.
PR	19-JUL-1999;	99US-0144335.	PR	14-OCT-1999;	99US-0159330.
PR	20-JUL-1999;	99US-0144352.	PR	18-OCT-1999;	99US-0159331.
PR	20-JUL-1999;	99US-0144632.	PR	21-OCT-1999;	99US-0159637.
PR	20-JUL-1999;	99US-0144884.	PR	21-OCT-1999;	99US-0159638.
PR	21-JUL-1999;	99US-0144814.	PR	21-OCT-1999;	99US-0159584.
PR	21-JUL-1999;	99US-0145086.	PR	21-OCT-1999;	99US-0160741.
PR	21-JUL-1999;	99US-0145088.	PR	21-OCT-1999;	99US-0160767.
PR	22-JUL-1999;	99US-0145085.	PR	22-OCT-1999;	99US-0160768.
PR	22-JUL-1999;	99US-0145087.	PR	22-OCT-1999;	99US-0160770.
PR	22-JUL-1999;	99US-0145089.	PR	25-OCT-1999;	99US-0160814.
PR	23-JUL-1999;	99US-0145192.	PR	25-OCT-1999;	99US-0160815.
PR	23-JUL-1999;	99US-0145145.	PR	26-OCT-1999;	99US-0160980.
PR	23-JUL-1999;	99US-0145218.	PR	26-OCT-1999;	99US-0160981.
PR	26-JUL-1999;	99US-0145224.	PR	26-OCT-1999;	99US-0160989.
PR	27-JUL-1999;	99US-0145276.	PR	28-OCT-1999;	99US-0161404.
PR	27-JUL-1999;	99US-0145913.	PR	28-OCT-1999;	99US-0161405.
PR	27-JUL-1999;	99US-0145918.	PR	29-OCT-1999;	99US-0161406.
PR	28-JUL-1999;	99US-0145919.	PR		99US-0161359.
PR	28-JUL-1999;	99US-0145951.	PR		99US-0161360.
PR	02-AUG-1999;	99US-0146386.	PR		99US-0161361.
PR	02-AUG-1999;	99US-0146388.	PR		99US-0161920.
PR	03-AUG-1999;	99US-0146389.	PR		99US-0161992.
PR	04-AUG-1999;	99US-0147038.	PR		99US-0161993.
PR	04-AUG-1999;	99US-0147204.	PR		99US-0162142.
PR	05-AUG-1999;	99US-0147302.			
PR	05-AUG-1999;	99US-0147260.			
PR	06-AUG-1999;	99US-0147303.			

Query Match   7.7%;   Score 179; DB 21;   Length 730;
Best Local Similarity   23.2%;   Pred. No. 0.00019;
Matches   115;   Conservative   52;   Mismatches 158;   Indels 170;   Gaps 24;

```
Qy    9  NTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAAT  68
            :|.|| :|:|::.:  | |:|  | ||:|:||
Db  132  stgakpgasgigsdgsslq-----sagtnpgadgtre------teknaggskpssgsagt 180

Qy   69  GAG----GNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGG 124
            |:|||||    ||||  |:||     ||  :|| || |:|
Db  181  npgasavgngete-knaggskpssgsagtnpga-----saggngeteknvggskpssg  232

Qy  125  QIGDNPLLKAMLLIARMMDGQSDQFGQPGT----GNNSASSGT------------  164
            :|.||         ::.||:::  :.:|||:
Db  233  kagtnp--------ganaggnggteknaggskssssgsaartnpgasaggngetv  277

Qy  165  ------SSSGGSPFNDLSGGKAPSGNYSPSGNYSPVSTFSPPSTPTSPSPLDFPSSPTK  217
                  :.: |:|::.: | |: |
Db  278  snigdtesnaggsksndganngasglesnag--stgtnfgaggtggigdtesdaggsktn 335

Qy  218  AAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAG  277
            : |.|:|:::||:. :   |:||:  :
Db  336  sgnggt--nd-----gasgigsndgstgtnpgag-------ggtdsnieg  371

Qy  278  SELGDGGQSENQKPLFILEDGAS-LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAIT  336
            :.| |||:  |     |||:  :|
Db  372  tennvggketn------pgasgign----sdgstgspeg---------tesnadg  408

Qy  337  VKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGN-W  395
            || : |: ::.:  | |:::: :    ||    | ||
Db  409  tktntggkesn----tgs--------esntnsspqkleaq-------ggnggnqw  444

Qy  396  DLNLSH--------ISAEDGKFSFVKSD--SEG----------LNVNTSDISL  428
            | |:|          :.|: |:.::|: |            :::|
Db  445  ddgtdhdgymklhavvggigieqirfdyvknqqlkegpfhgvkgrggtsteishpdeyl  504

Qy  429  GDVENHYKVPMSANL  443
            || :||
Db  505  vsveglyd--ssni  516
```

RESULT 12

AAG29581

```
ID   AAG29581 standard; Protein; 752 AA.
XX
AC   AAG29581;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana protein fragment SEQ ID NO: 35221.
XX
KW   Protein identification; signal transduction pathway; metabolic pathway;
KW   hybridisation assay; genetic mapping; gene expression control; promoter;
KW   termination sequence.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999; 99US-0121825.
PR   25-MAR-1999; 99US-0123180.
PR   09-MAR-1999; 99US-0123548.
PR   23-MAR-1999; 99US-0125788.
PR   25-MAR-1999; 99US-0126264.
PR   29-MAR-1999; 99US-0126785.
PR   01-APR-1999; 99US-0127462.
PR   06-APR-1999; 99US-0128234.
PR   08-APR-1999; 99US-0128714.
PR   16-APR-1999; 99US-0129845.
PR   19-APR-1999; 99US-0130077.
PR   21-APR-1999; 99US-0130449.
PR   23-APR-1999; 99US-0130510.
PR   23-APR-1999; 99US-0130891.
PR   28-APR-1999; 99US-0133149.
PR   30-APR-1999; 99US-0132048.
PR   30-APR-1999; 99US-0132407.
PR   04-MAY-1999; 99US-0132484.
PR   05-MAY-1999; 99US-0132485.
PR   06-MAY-1999; 99US-0132486.
PR   06-MAY-1999; 99US-0132487.
PR   07-MAY-1999; 99US-0132863.
PR   11-MAY-1999; 99US-0134256.
PR   14-MAY-1999; 99US-0134218.
PR   14-MAY-1999; 99US-0134219.
PR   14-MAY-1999; 99US-0134221.
PR   14-MAY-1999; 99US-0134370.
PR   18-MAY-1999; 99US-0134768.
PR   19-MAY-1999; 99US-0134941.
PR   20-MAY-1999; 99US-0135124.
PR   21-MAY-1999; 99US-0135353.
PR   24-MAY-1999; 99US-0135629.
PR   25-MAY-1999; 99US-0136021.
PR   27-MAY-1999; 99US-0136392.
PR   28-MAY-1999; 99US-0136782.
PR   01-JUN-1999; 99US-0137222.
PR   03-JUN-1999; 99US-0137528.
PR   04-JUN-1999; 99US-0137502.
PR   07-JUN-1999; 99US-0137724.
PR   08-JUN-1999; 99US-0138094.
PR   10-JUN-1999; 99US-0138540.
PR   10-JUN-1999; 99US-0138847.
PR   14-JUN-1999; 99US-0139119.
PR   16-JUN-1999; 99US-0139452.
PR   16-JUN-1999; 99US-0139453.
PR   17-JUN-1999; 99US-0139492.
PR   18-JUN-1999; 99US-0139454.
PR   18-JUN-1999; 99US-0139455.
PR   18-JUN-1999; 99US-0139456.
PR   18-JUN-1999; 99US-0139457.
PR   18-JUN-1999; 99US-0139458.
PR   18-JUN-1999; 99US-0139459.
PR   18-JUN-1999; 99US-0139460.
PR   18-JUN-1999; 99US-0139461.
PR   18-JUN-1999; 99US-0139462.
PR   18-JUN-1999; 99US-0139463.
PR   18-JUN-1999; 99US-0139750.
PR   21-JUN-1999; 99US-0139763.
PR   22-JUN-1999; 99US-0139817.
PR   23-JUN-1999; 99US-0139899.
PR   23-JUN-1999; 99US-0140353.
PR   23-JUN-1999; 99US-0140354.
PR   24-JUN-1999; 99US-0140695.
PR   28-JUN-1999; 99US-0140823.
PR   29-JUN-1999; 99US-0140991.
PR   30-JUN-1999; 99US-0141287.
PR   01-JUL-1999; 99US-0141842.
PR   01-JUL-1999; 99US-0142154.
PR   02-JUL-1999; 99US-0142055.
PR   06-JUL-1999; 99US-0142390.
PR   08-JUL-1999; 99US-0142803.
PR   09-JUL-1999; 99US-0142920.
PR   12-JUL-1999; 99US-0142977.
PR   13-JUL-1999; 99US-0143542.
PR   14-JUL-1999; 99US-0143624.
PR   15-JUL-1999; 99US-0144005.
PR   16-JUL-1999; 99US-0144085.
PR   16-JUL-1999; 99US-0144086.
PR   19-JUL-1999; 99US-0144325.
PR   19-JUL-1999; 99US-0144331.
PR   19-JUL-1999; 99US-0144332.
PR   19-JUL-1999; 99US-0144333.
PR   19-JUL-1999; 99US-0144334.
PR   19-JUL-1999; 99US-0144335.
PR   20-JUL-1999; 99US-0144352.
PR   20-JUL-1999; 99US-0144632.
```

```
PR   20-JUL-1999;    99US-0144884.
PR   21-JUL-1999;    99US-0144814.
PR   21-JUL-1999;    99US-0145086.
PR   21-JUL-1999;    99US-0145088.
PR   22-JUL-1999;    99US-0145085.
PR   22-JUL-1999;    99US-0145087.
PR   22-JUL-1999;    99US-0145089.
PR   22-JUL-1999;    99US-0145192.
PR   23-JUL-1999;    99US-0145145.
PR   23-JUL-1999;    99US-0145218.
PR   23-JUL-1999;    99US-0145224.
PR   26-JUL-1999;    99US-0145276.
PR   27-JUL-1999;    99US-0145913.
PR   27-JUL-1999;    99US-0145918.
PR   27-JUL-1999;    99US-0145919.
PR   28-JUL-1999;    99US-0145951.
PR   02-AUG-1999;    99US-0146386.
PR   02-AUG-1999;    99US-0146388.
PR   02-AUG-1999;    99US-0146389.
PR   03-AUG-1999;    99US-0147038.
PR   04-AUG-1999;    99US-0147204.
PR   04-AUG-1999;    99US-0147302.
PR   05-AUG-1999;    99US-0147192.
PR   05-AUG-1999;    99US-0147260.
PR   06-AUG-1999;    99US-0147303.
PR   06-AUG-1999;    99US-0147416.
PR   09-AUG-1999;    99US-0147493.
PR   09-AUG-1999;    99US-0147935.
PR   10-AUG-1999;    99US-0148171.
PR   11-AUG-1999;    99US-0148319.
PR   12-AUG-1999;    99US-0148341.
PR   13-AUG-1999;    99US-0148565.
PR   13-AUG-1999;    99US-0148684.
PR   16-AUG-1999;    99US-0149368.
PR   17-AUG-1999;    99US-0149175.
PR   18-AUG-1999;    99US-0149426.
PR   20-AUG-1999;    99US-0149722.
PR   20-AUG-1999;    99US-0149723.
PR   20-AUG-1999;    99US-0149929.
PR   23-AUG-1999;    99US-0149902.
PR   25-AUG-1999;    99US-0149930.
PR   26-AUG-1999;    99US-0150566.
PR   26-AUG-1999;    99US-0150884.
PR   27-AUG-1999;    99US-0151065.
PR   27-AUG-1999;    99US-0151066.
PR   27-AUG-1999;    99US-0151080.
PR   30-AUG-1999;    99US-0151303.
PR   31-AUG-1999;    99US-0151438.
PR   01-SEP-1999;    99US-0151930.
PR   07-SEP-1999;    99US-0152363.
PR   10-SEP-1999;    99US-0153070.
PR   13-SEP-1999;    99US-0153758.
PR   15-SEP-1999;    99US-0154018.
PR   16-SEP-1999;    99US-0154039.
PR   20-SEP-1999;    99US-0154779.
PR   22-SEP-1999;    99US-0155139.
PR   23-SEP-1999;    99US-0155486.
PR   24-SEP-1999;    99US-0155659.
PR   28-SEP-1999;    99US-0156458.
PR   29-SEP-1999;    99US-0156596.
PR   04-OCT-1999;    99US-0157117.
PR   05-OCT-1999;    99US-0157753.
PR   06-OCT-1999;    99US-0157865.
PR   07-OCT-1999;    99US-0158029.
PR   08-OCT-1999;    99US-0158232.
PR   12-OCT-1999;    99US-0158369.
PR   13-OCT-1999;    99US-0159293.
PR   13-OCT-1999;    99US-0159294.
PR   13-OCT-1999;    99US-0159295.
PR   14-OCT-1999;    99US-0159329.
PR   14-OCT-1999;    99US-0159330.
PR   14-OCT-1999;    99US-0159331.
PR   14-OCT-1999;    99US-0159637.
```

```
PR   14-OCT-1999;    99US-0159638.
PR   16-OCT-1999;    99US-0159584.
PR   21-OCT-1999;    99US-0160741.
PR   21-OCT-1999;    99US-0160767.
PR   21-OCT-1999;    99US-0160768.
PR   21-OCT-1999;    99US-0160770.
PR   21-OCT-1999;    99US-0160814.
PR   22-OCT-1999;    99US-0160815.
PR   22-OCT-1999;    99US-0160980.
PR   22-OCT-1999;    99US-0160981.
PR   22-OCT-1999;    99US-0160989.
PR   25-OCT-1999;    99US-0161404.
PR   25-OCT-1999;    99US-0161405.
PR   26-OCT-1999;    99US-0161406.
PR   26-OCT-1999;    99US-0161359.
PR   26-OCT-1999;    99US-0161360.
PR   26-OCT-1999;    99US-0161361.
PR   28-OCT-1999;    99US-0161920.
PR   28-OCT-1999;    99US-0161992.
PR   28-OCT-1999;    99US-0161993.
PR   29-OCT-1999;    99US-0162142.
```

```
Query Match             7.7%;    Score 179;  DB 21;  Length 752;
Best Local Similarity   23.2%;   Pred. No. 0.0002;
Matches  115;    Conservative   52;   Mismatches  158;   Indels   170;   Gaps   24;

Qy      9 NTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAAT 68
          ::|:|    |  |:|:      |        |        ||   |:||||   | |||
Db    154 stgakpgasglgsdsgslg------sagtnpgadgtre----teknaggskpssgssagt 202

Qy     69 GAG----GNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGG 124
          | |     | | |||::::|||:|  ::|||::|         |
Db    203 npgasavgngete--knaggskpssgsagtnpga-----saggngeteknvggskpssg 254

Qy    125 QIGDNPLLKAMLKLIARMMDGQSDQFQPGT----GNNSASSGT----------------- 164
          :  :||                           ||   | |
Db    255 kagtnp------------------ganagnggteknaggskssgsartnpgasaggngetv 299

Qy    165 ------SSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTK 217
          :::||||||      | |:|||     |  |
Db    300 snigdtesnaggsksndganngasglesnag--stgtnfgaggtggigdtesdaggsktn 357

Qy    218 AAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGGVFDGKGQTFTAG 277
          |::||:      |: :  ||:
Db    358 sgnggt--nd----gasglgsndgstnpgag-----------------ggtdsnleg 393

Qy    278 SELGDGGQSENQKPLFILEDGAS-LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAIT 336
          :  :||    ||   ||     ::
Db    394 tennvggketn------pgasgign----sdgstgtspeg-----------tesnadg 430

Qy    337 VKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGN-W 395
          |: : |:|:      ::|:                 |:|:
Db    431 tktntggkesn------tgs----------esntnsspqkleaq-------ggnggnqw 466

Qy    396 DLNLSH---------------ISAEDGKFSFVKSD--SEG---------------LNVNTSDISL 428
          |  |:                | |:                              ||
Db    467 ddgtdhdgvmklhvavgglgleqirfdyvkngglkegpfhgvkgrggtstielshpdeyl 526

Qy    429 GDVENHYKVPMSANL 443
          || |:  ::|:
Db    527 vsveglyd--ssnl 538
```

```
RESULT 13
AAW56319
ID   AAW56319 standard; Protein; 2042 AA.
XX
AC   AAW56319;
XX
DT   19-AUG-1998  (first entry)
XX
DE   Haemophilus paragallinarum antigenic protein #1.
XX
```

```
KW   Haemophilus paragallinarum; antigenic protein; HI antibody; diagnosis;
KW   vaccine; chicken infectious coryza; CIC; fowl.
XX
OS   Haemophilus paragallinarum.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..70
FT                  /label= signal
FT   Protein        71..2042
FT                  /note= "antigenic protein"
XX
PN   WO9812331-A1.
XX
PD   26-MAR-1998.
XX
PF   12-SEP-1997;   97WO-JP03222.
XX
PR   19-SEP-1996;   96JP-0271408.
XX
PA   (KAGA ) ZH KAGAKU & KESSEI RYOHO KENKYUSHO.
PA   (KAGA ) CHEMO-SERO-THERAPEUTIC RES INST.
XX
PI   Hamada F, Matsuo K, Sakaguchi M, Tokiyoshi S, Tokunaga E;
XX
DR   WPI; 1998-230318/20.
DR   N-PSDB; AAV22834.
XX
PT   Antigenic polypeptide from Haemophilus paragallinarum induces HI
PT   antibody production - and is useful for diagnosis of and preparation
PT   of vaccines for chicken infectious coryza
XX
PS   Claim 1; Page 51-69; 108pp; Japanese.
XX
CC   The present sequence represents an antigenic protein derived from
CC   Haemophilus paragallinarum strain A-221. The antigenic protein
CC   stimulates the production of HI antibodies in fowl. The protein
CC   and DNA coding for it can be used in the preparation of vaccines
CC   for the prevention of chicken infectious coryza (CIC). The protein
CC   and its antibodies can be used in the diagnosis and treatment of CIC.
XX
SQ   Sequence  2042 AA;

Query Match             7.5%;    Score 174;   DB 19;   Length 2042;
Best Local Similarity  23.2%;    Pred. No. 0.0016;
Matches  114;  Conservative   64;  Mismatches  189;   Indels 124;   Gaps  22;

Qy    29 NANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGGNDQTTGVGNAGGLNGR 88
          ||::| |  :||| |  :|    |     |          :| |||    | : |||
Db   905 nasnadsarnvggltpqgaiaqlkan-nnylndgakgqdsiafgwqaktsgann--glagk 961

Qy    89 KGTA-GTTPQSDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMMDGQS 147
          : | | |:|:|||| :| ||  |||:|    | ||||||              | |
Db   962 qalalgfqanssaenais-igtns-dtsmt----gavalgkgatvtaggk------ps 1007

Qy   148 DQFGQPGTGNNSASSGTSSS--GGSPFNDLSGGKAPSG----------NSPSGNYS 191
          | |    |  |  |  | ||| ::| | |||          |:|| | |:  |  |
Db  1008 lalggdstvansaisrtsspminglifnnfagspetlgvlsigtagrerkivnvaagdvs 1067

Qy   192 PVSTFSPPSTPTSPTS-PLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTS-AG 249
          | :  :  | | | |                       |:     |
Db  1068 qastealngsqlyatnfmlskvaqsvksnfg--------gnvlgtdgtitftnigg 1116

Qy   250 ANQTVLHDTIT------------VKAGQ--------VFDGKGQTFTAGSELGDGG-------- 284
          |  | | :| |         |||::          :|||:  : |: : | :  :
Db  1117 tgqatihdinnvltkgjylkadqndptgnqgkvelgnaitlsatnqwanngvnyktnn 1176

Qy   285 ---QSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDN--------LHVTNV 330
          :| | | :|     :| |:           |  || |       | |:::
Db  1177 lttynsqngtilgmredpsvkqitagt-----ynttgdannknqlnntlqgttleatgi 1231

Qy   331 ----------GEDAITVKFPNSAG--KKSHVEITNSSFEHASDKIILQLNADTNLSVD 374
          ||| ||   | | :||| :  | :| |||:   :
Db  1232 tssvgstnyagfslgadsvfskggagtvklsgvsdatatdaatlkqvkeyrttlvgdn 1291

Qy   375 NVKAKDFGTFVRTNGGQQG---NWDLNLSHISAEDGKFSFVKSDSEGL-------NVN 422
          :| :|:|: :  |:  : :|    | :  :||  : : |:     ||        :|:
Db  1292 ditaad-----rsggtsngitynlslnkgtvsateekvvsgktvyeairnaltgniftig 1346

Qy   423 TSDISLGDVEN 433
          |::|:|  :|
Db  1347 lddtlnkinn 1357
```

RESULT 14
AAW06600
```
ID   AAW06600 standard;  Protein; 344 AA.
XX
AC   AAW06600;
XX
DT   30-MAR-1997  (first entry)
XX
DE   Hypersensitive response elicitor protein.
XX
KW   Hypersensitive response; elicitor; Pseudomonas solanacearum; plant;
KW   disease-resistance; Escherichia coli; infiltration; virus;
KW   bacterium; fungus; pathogen; biological control agent.
XX
OS   Pseudomonas solanacearum.
XX
PN   WO9639802-A1.
XX
PD   19-DEC-1996.
XX
PF   05-JUN-1996;   96WO-US08819.
XX
PR   07-JUN-1995;   95US-0475775.
XX
PA   (CORR ) CORNELL RES FOUND INC.
XX
PI   Beer SV, Wei Z;
XX
DR   WPI; 1997-051614/05.
DR   N-PSDB; AAT49316.
XX
PT   Imparting pathogen resistance to plants - with hypersensitive
PT   response elicitor polypeptide or protein
XX
PS   Claim 13; Page 50-51; 69pp; English.
XX
CC   This sequence represents a hypersensitive response elicitor from
CC   Pseudomonas solanacearum. The elicitor may be used in a new method
CC   for imparting pathogen resistance to plants, by application of the
CC   elicitor in a non-infectious form to plant cells, by spraying,
CC   injection, leaf abrasion, or plant infection with recombinant
CC   bacteria (non-infectious to the host plant, e.g. Escherichia coli)
CC   expressing the elicitor as a biological control agent, to alllow
CC   recombinant protein infiltration into the plant. The method
CC   confers virus, bacterium or fungus disease-resistance on crops and
CC   ornamental plants.
XX
SQ   Sequence  344 AA;

Query Match             7.2%;    Score 165.5;  DB 18;   Length 344;
Best Local Similarity  32.0%;    Pred. No. 0.00064;
Matches  57;  Conservative   20;  Mismatches  80;   Indels  21;   Gaps  7;

Qy    18 QSGGD--NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN 73
          |:|||  :||:| ||||||| ::| |||||:|| | : ||  | |::|| |  |: |||
Db   134 qpggndkgngvggangakgagggggglaealqeleqllaqigggagggaggvggagga 193

Qy    74 DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGQIGDN 129
          : :| ||| |   | : | ||     | ||||::|| ||| |  |  | :  |||:|| |
Db   194 dggsgaggaggangadggngvngngangpqnagdvngangad-----dgsedqggltgvl 248
```

Qy   130 PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP--FNDLSGGK 179
              :| :|| : :|| | ||| : ||| : || :   :| :||:||
Db   249 qklmkilnalvqmmqgglgggnqaggskgagnaspasganpganqpgsaddqssgq 306

Db   249 qklmkilnalvqmmqgglgggnqaggskgagnaspasganpganqpgsaddqssgq 306

Search completed: October 18, 2001, 09:10:01
Job time: 76 sec

RESULT 15
AAW75865
ID AAW75865 standard; Protein; 344 AA.
XX
AC AAW75865;
XX
DT 07-DEC-1998 (first entry)
XX
DE Pseudomonas solanacearum hypersensitive response elicitor (HRE).
XX
KW Hypersensitive response elicitor; HRE; insect resistance;
KW biological control; transgenic plant.
XX
OS Pseudomonas solanacearum.
XX
PN WO9837752-A1.
XX
PD 03-SEP-1998.
XX
PF 26-FEB-1998;   98WO-US03604.
XX
PR 28-FEB-1997;   97US-0039226.
XX
PA (CORR) CORNELL RES FOUND INC.
XX
PI Wei Z, Zitter TA;
XX
DR <-WPI; 1998-495374/42.
DR N-PSDB; AAV54609.
XX
PT Use of hypersensitive response elicitor polypeptide - for
PT application to plants or seeds or transgenic plants or seeds for the
PT control of insects.
XX
PS Disclosure; Page 14-15; 75pp; English.
XX
CC This is the amino acid sequence of a hypersensitive response
CC elicitor (HRE) of Pseudomonas solanacearum. The invention
CC relates to the use of a HRE polypeptide or protein to control
CC insects on plants or plants grown from seed treated with HRE.
CC Also claimed is a method of insect control for plants that involves:
CC (a) providing a transgenic plant or seed transformed with a DNA
CC molecule (see AAV54606-09) encoding a HRE polypeptide or protein (see
CC AAW75862-67); and (b) growing the transgenic plants or transgenic
CC plants produced from the transgenic seeds to control insects. HRE
CC prevents direct insect damage to plants by feeding injury. It kills
CC insects close to plants, and interferes with insect larval feeding
CC on such plants. It also prevents insects from colonising host
CC plants and releasing phytotoxins which result in disease damage to
CC plants.
XX
SQ Sequence   344 AA;

Query Match   7.2%;   Score 165.5; DB 19; Length 344;
Best Local Similarity   32.0%; Pred. No. 0.00064;
Matches   57; Conservative   20; Mismatches   80; Indels   21; Gaps   7;

Qy   18 QSGGD---NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN 73
              | :: ||:||| | | : :: : ||:|:   : :| ||||| ||||
Db   134 qpggndkgngvvgangakgagggggglaealqeleqilaqlgggggaggaggvggagga 193

Qy   74 DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGGQIGDN 129
              | | |:| | | || || ||| : || :| :| | ||| || | ||
Db   194 dgggsagggaggangadggngvngngangpqnagdvngangad----dgsedqgg1tgv1 248

Qy   130 PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP--FNDLSCGK 179
              :| :|| : :|| | ||| : ||| : || :   :| :||:||

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 18, 2001, 09:08:45 ; Search time 17.53 Seconds
(without alignments)
1942.386 Million cell updates/sec

Title: US-09-596-958-2
Perfect score: 2310
Sequence: 1 MSILTLNNTSSSPGLFQSG..........LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Database :   PIR_68:*
                 1: pir1:*
                 2: pir2:*
                 3: pir3:*
                 4: pir4:*

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2310	100.0	447	2	T18447	HrpW protein - Erw
2	353.5	15.3	221	2	A70045	pectate lyase homo
3	233.5	10.1	219	2	S68364	pectate lyase (EC
4	207	9.0	1660	2	A70869	hypothetical glyci
5	199.5	8.6	1306	2	A70934	hypothetical glyci
6	193	8.4	694	2	F70868	hypothetical glyci
7	188	8.1	639	2	D70931	hypothetical glyci
8	186.5	8.1	591	2	B70523	hypothetical glyci
9	186	8.1	1329	2	E70917	hypothetical glyci
10	185.5	8.0	1381	2	E70806	hypothetical glyci
11	185.5	8.0	1489	2	D70807	hypothetical glyci
12	183.5	7.9	242	2	A45724	pectate lyase (EC
13	183	7.9	588	2	F70971	hypothetical glyci
14	182	7.9	606	2	H70816	hypothetical glyci
15	182	7.9	731	2	C70974	hypothetical glyci
16	181.5	7.9	484	2	G70846	hypothetical glyci
17	180.5	7.8	1079	2	B70807	hypothetical glyci
18	179	7.7	730	2	F96559	hypothetical prote
19	179	7.7	1538	2	H70846	hypothetical glyci
20	179	7.7	3016	2	S77300	hypothetical prote
21	178	7.7	741	2	G70917	hypothetical glyci
22	177.5	7.7	1901	2	F70806	hypothetical glyci
23	177.5	7.7	2232	2	T34434	hypothetical glyci
24	176.5	7.6	562	2	B70953	hypothetical glyci
25	175	7.6	645	2	F70825	probable PPE prote
26	174	7.5	615	2	H70589	hypothetical glyci
27	172.5	7.5	584	2	A70804	hypothetical glyci
28	172	7.4	3507	2	T34513	hypothetical prote
29	171.5	7.4	667	2	A70893	hypothetical glyci
30	171.5	7.4	1778	2	T50074	probable nucleopor
31	171	7.4	463	2	B70893	hypothetical glyci
32	169.5	7.3	914	2	H70987	hypothetical glyci
33	169	7.3	525	2	D70878	hypothetical glyci
34	168.5	7.3	398	2	OZZQAS	circumsporozoite p
35	167.5	7.2	1011	2	F70620	hypothetical glyci
36	166.5	7.2	532	2	F70580	hypothetical glyci
37	166.5	7.2	714	2	A70807	hypothetical glyci
38	166	7.2	461	2	F70571	hypothetical glyci
39	166	7.2	603	2	A70770	hypothetical glyci
40	166	7.2	1207	2	T23754	hypothetical prote
41	165.5	7.2	344	2	S41707	PopA1 protein - Ps
42	165	7.1	591	2	S74999	iron-regulated pro
43	165	7.1	749	2	A70812	hypothetical glyci
44	165	7.1	959	2	B44402	nuclear pore compl
45	164	7.1	767	2	E70895	hypothetical glyci

ALIGNMENTS

RESULT 1
T18447
HrpW protein - Erwinia amylovora
C;Species: Erwinia amylovora
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18447
R;Gaudriault, S.
submitted to the EMBL Data Library, May 1998
A;Reference number: -218936
A;Accession: T18447
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-447 <GAU>
A;Cross-references: EMBL:Y13831; NID:e1293450; PID:e1293451; PIDN:CAA74158.1
A;Experimental source: strain CFBP1430; specific host Pommoideae
C;Genetics:
A;Note: hrpW

```
Query Match             100.0%;  Score 2310;  DB 2;  Length 447;
Best Local Similarity   100.0%;
Matches  447;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MSILTLNNTSSSPGLFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSILTLNNTSSSPGLFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60

Qy   61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG 120

Qy  121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180

Qy  181 PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240

Qy  241 NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300

Qy  301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360

Qy  361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420

Qy  421 VNTSDISLGDVENHYKVPMSANLKVAE 447
```

```
                         |||||||||||||||||||||||||||
Db    421 VNFSDISLGDVENHYKVPMSANLKVAE 447

RESULT  2
A70045
pectate lyase homolog yvpA - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 15-Oct-1999
C;Accession: A70045
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033
A;Accession: A70045
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-221 <KUN>
A;Cross-references: GB:Z99121; GB:AL009126; NID:g2635827; PIDN:CAB15500.1; PID:e1186183;
A;Experimental source: strain 168
C;Genetics:
A;Gene: yvpA

Query Match      15.3%;     Score 353.5;  DB 2;    Length 221;
Best Local Similarity  41.8%;       Pred. No. 5.6e-13;
Matches  76;  Conservative  31;  Mismatches  70;  Indels   5;  Gaps   1;

Qy  250 ANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDD 309
            |  |:||: ||||||||||||||||||| |: |:|:||:||||| ::
Db   27 AADKVVHETIIVPKNTTYDGKGQRFVAGKELGDGSQSENQDPVFRVEDGATLKNVVLGAP  86

Qy  310 GADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADT 369
            : |||: ||:  |||||||| : :||||| |||||||:|
Db   87 AADGVHTYGNVNIQNVKWEDVGEDALTVK-----KEGKVTIDGGSAQKASDKIFQINKAS 141

Qy  370 NLSVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLG 429
            :|: |||||||||:|::  | | :||  :: ::: |:  ::| :
Db  142 TFTVKNFTADNGGKFIRQLGGSTFHVDVIIDKCTITNMKEALFRTDSKTSVRMTNTRYS 201

Qy  430 DV 431
            :|
Db  202 NV 203
```

```
RESULT  3
S68364
pectate lyase (EC 4.2.2.2) C - fungus (Fusarium solani)
C;Species: Fusarium solani
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
C;Accession: S68364
R;Guo, W.; Gonzalez-Candelas, L.; Kolattukudy, P.E.
Arch. Biochem. Biophys. 323, 352-360, 1995
A;Title: Cloning of a new pectate lyase gene pelC from Fusarium solani f. sp. pisi (Nect
A;Reference number: S68364; MUID:96063610
A;Accession: S68364
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <GUO>
A;Cross-references: EMBL:U13049
C;Genetics:
```

```
A;Gene: pelC
A;Introns: 52/1; 102/3
C;Keywords: carbon-oxygen lyase

Query Match      10.1%;     Score 233.5;  DB 2;    Length 219;
Best Local Similarity  30.1%;       Pred. No. 2.3e-06;
Matches  63;  Conservative  39;  Mismatches  72;  Indels  35;  Gaps   6;

Qy  220 GGSTPVTDHPDPVGSAGICAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSE 279
            || ||                         |||||:|||:||:|||||:|||:|||:
Db    8 GGVPKPTDH------------------ISNS----KVIEVKAGQVDGKWAKYDRGSG  43

Qy  280 LGDG-GQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVK 338
            :      :: : : |:|||||||| | |: : | :
Db   44 ACKGQNEGGDKDAVFLLHEGATLKNVIIGKDQSEGVHCKGHCTLEFWFEDVCEDAISIK 103

Qy  339 PNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRT--NGGQGNWD 396
            ::|||:  ::      :: :|  |||  :  ::|  : |||||  :|:| :|||||||
Db  104 EDRAGKESW--IIGGGAYHASDKVVQHNGCGTVNIINFYVEDYGKLYRSCGNCSKQCKRN 161

Qy  397 LNLSHISAEDGKFSFVKSDSEGLNVNTSD 425
            ||  :  |:|:  :|  : |:|:
Db  162 VYIEGVTAKNG-----GELAGINANYGD 184
```

```
RESULT  4
A70869
hypothetical glycine-rich protein Rv2490c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70869
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete gen
A;Reference number: A70500; MUID:98295987
A;Accession: A70869
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1660 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16067.1; PID:g270
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2490c
C;Superfamily: collagen alpha 1(IV) chain

Query Match       9.0%;     Score 207;  DB 2;    Length 1660;
Best Local Similarity  28.1%;       Pred. No. 0.00077;
Matches  91;  Conservative  19;  Mismatches  130;  Indels  84;  Gaps  15;

Qy    7 NNNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSP--QSG  64
           :  |  |:     :  |                        |
Db  476 NGGAGGNGGLV--GDGGAGGHGGDGAG----------------AGYADMTAIFLGSSGTPGEDGG 523

Qy   65 N----AATGAGG-NDQTTGVGNAGGLNGRKGTAGTTEQSDSQNMLSEMGNNGLDQA--IT 117
            |                  | |:|              :|:||:|||     |:
Db  524 NGGAGGAGGAHAGDGGAGGNGGAGGAGGAGGNGAHGFNAVLVSDGGNGGDGGAGGRG 583

Qy  118 PDGQGGQGIGDNPLLKAMLKLIARMMDGSDQFGQPGTGNNSASSGTSSSGG---SPFND 174
            :   :|| ::|  |     ::| : |
Db  584 GDGGAGGAGGDAPAGRA-------GSQGVGGDGGAGGAGGAPGNGGSGGRGDMAFKD 633

Qy  175 LSGGKAPSGNSP-SGNYSPVSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPVTDHPDPVG 233
            :  : |:  |:  |      :             :
Db  634 GDGGAGGDGGDPGAGGKGAG-------------------GAGATEGVTGATGATVHSG--G 674

Qy  234 SAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLF 293
            :  :||                :||  | |                |:
Db  675 NGGKG-GNGADATVAGAN-------------------------GGKGGAGGNGGLVGDGGAG------ 710
```

Qy 294 ILEDG----ASLKNVTMGDDGADG 313
              :|:|||||
Db 711 --GDGGSGAAGANGANVGEDGADG 732

RESULT 5
A70934
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: A70934
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1306 <COL>
A;Cross-references: GB:AL021942; GB:AL123456; NID:g3242298; PIDN:CAA17449.1; PID:g290963
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0578c
C;Superfamily: collagen alpha 1(IV) chain

Query Match   8.6%; Score 199.5; DB 2; Length 1306;
Best Local Similarity 26.0%; Pred. No. 0.0015;
Matches   88; Conservative 16; Mismatches 122; Indels 113; Gaps 11;

Qy   8 NNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAA 67
             |:|||| :|||| :| |||||||||||||
Db 754 NGVAGSQGPGAGGDGGTGGVGVGGNGGRGIDGADGAT-------------------- 789

Qy   68 TGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGGIG 127
             || ||| ||| ||||
Db 790 --AGARGQDGGAGGAGGAGGKGGRGTGGP----------------GGAGPAGTTGSQGAGNG 832

Qy 128 DNPLLKAMLKLIARMMDGQSDQFQQPGTGNNSASSGTSSSGGSPFNDLSGGKA-PSGNSP 186
             | | || ||: |:| ||| ||||||| ||| |||
Db 833 --------------GSGGTGGDPGDGGNGANGSVFTNNGIGGNGGNGNAGPSGAGG 875

Qy 187 SGNYSPVSTFSPPSTPTSPTSPL------------DFPSSPTKAAG----------G 221
             ||| ||| ||||||:|||||
Db 876 SGGAG--STFG---ATGSSSSIHVNGGNGGNGGNGDHALSGNGAAGGNGGNGGNGSLRG 929

Qy 222 STPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDCKGQTFTAGSELG 281
             |||:| |||||
Db 930 SGGAGGHGGNGGNASRGMGGDGGTGGAGGN---------:|:----AQIGNG-----GAGGNGG 973

Qy 282 DGGQSENQKPLFIL------EDGASLKNVTMGDDGADG 313
             | ||| :: :||||
Db 974 DGGTGSDGNPGAITGSGGRGGDGVGGQGGSVAGDGADG 1012

RESULT 6
F70868
hypothetical glycine-rich protein Rv2487c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon,
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: F70868

A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-694 <COL>
A;Cross-references: GB:AL021246; GB:AL123456; NID:g3261507; PIDN:CAA16064.1; PID:g279
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv2487c
C;Superfamily: elastin

Query Match   8.4%; Score 193; DB 2; Length 694;
Best Local Similarity 25.4%; Pred. No. 0.0016;
Matches   86; Conservative 28; Mismatches 140; Indels 84; Gaps 14;

Qy   18 QSGGDNGLGGH-NANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGNDQT 76
             |:||| |||| ||||| |||||
Db 256 QTGGDGGTGGHGTAGTPTGGTGGDAT----------------ATAGSGKATGGAGDGGT 300

Qy   77 TGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGQICDNPLLKAML 136
             | || || | | |||
Db 301 AAAGGGGGNGGDGVA----QGDIASAFGDGGNGSDGVAAGSGGSGGAGG----AFV 352

Qy 137 KLIARMMDGQSDQFQQPGTGNNSASSGT-SSSGGSPFNDLSGGK--APSGNSPSGNYSPV 193
             :||| ||| :|| | |||||||
Db 353 HIATATSTGGSGGFG--GNGAASAASGADGGAGNGGAGGLLFGDGGNGGAGGAGGI 410

Qy 194 STFSPPSTPTSPTSPLDFPSSPTKAAGGSTPVT-DHPDP------------VGSAG 236
             | :|:   :|:|
Db 411 GG-------DGATGGPGGSGGNAGIARFDSPDPEAEPDVVGGKGGDGGKGGSG 456

Qy 237 IGAGNSVAFTSAGANQTVLHDTITVKAGQVFDCKGQTFTAGSELGDGGQSENQKPLFILE 296
             :|:| ||| ||| |||
Db 457 LGVGAGGTGGAGGNG---------GAGGLLFGNGGN--GGNAGAGG-------D 493

Qy 297 DGASLKNVTMGDDGADGIHLYGDAKIDNL-HVTNVGED 333
             | :||| |||| ||: |
Db 494 GGAGVAGGVGNGGGGGTATFHEDPVAGWAVGGVGGD 531

RESULT 7
D70931
hypothetical glycine-rich protein Rv1803c - Mycobacterium tuberculosis (strain H37Rv)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 24-Nov-1999
C;Accession: D70931
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500: MUID:98295987
A;Accession: D70931
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-639 <COL>
A;Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAA17724.1; PID:e125
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1803c
C;Superfamily: unassigned collagens

Query Match   8.1%; Score 188; DB 2; Length 639;
Best Local Similarity 22.9%; Pred. No. 0.0027;
Matches   96; Conservative 33; Mismatches 144; Indels 146; Gaps 17;

Qy   7 NNNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNA 66
             | ||:|| ||| ||||
Db 159 NGGNGGSAGLWGSGNGGQGGGAGANGAAGQP----------------GK 191

Qy   67 ATGAGGNDQT----TCVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQG 122
             | |:|||

```
Db    354  GAGGNGGHVSGGSVNT---------AGA--GGKGGNGGTGGGAGGPGGH---------- 390

Qy    298  GASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVG 331
           || ::| |||||||| |||||| |||||||||||
Db    391  GGSVLSGPVGDSGNGGAGGDGGAGVSATDIAGTG 424

RESULT  9
E70917
hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70917
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Skelton, K.; Seeger, K.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
A;Accession: E70917
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1329 <COL>
A;Cross-references: GB:Z95844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g21310
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1450c
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo

Query Match       8.1%;   Score 186;   DB 2;   Length 1329;
Best Local Similarity   25.2%;   Pred. No. 0.0084;
Matches   86;  Conservative   25;  Mismatches 126;  Indels 104;  Gaps  14;

Qy      7  NNNTSSSPGLFQSGGDNGLGGHNANSALGGQP---------IDRQTIEQMAQLLAELLKS 57
                                                     ::|
Db    704  NGANATTPG-AKGGDGHGGPGAQGGNGGQGGPGGLAGNLFGQNGIQGVGGSGGKGGAG 761

Qy     58  LLSPQSGNAAT-----------------GAGGNDQTTGVGNAGGLN--GRKGTAGTTPQSDSQ 101
           |:  ||| |||                 ::| | |||| ||| |||  |||
Db    762  GLAGDGNGANGNFAFGDNGHGGNGGNGGNPGAGGQGGGGGAGSTPGAKGAHGFTPTSGGD 821

Qy    102  NMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMMDGQSDOFGQPGTGNNSAS 161
                                          | |||
Db    822  G--GDGNGNGNSQVVGGNGGDGGNGNG-------------------GSAGTGGNGGR 858

Qy    162  SGTSSSGGSPFNDLSGGK-APSGNSPSGNYSPVSTFSPPSTPSTPSTPLDFPSSPTKAAG 220
           :|                 |::|
Db    859  GGDGAFGGMSANATNPGENGPNGN-PGGN------------------------GGAG 890

Qy    221  GSTPVTDHPDPVGSAGI----GAGNSVAFTSAGANQTVLHDTITVKA-GQV--FDGKGQ 272
           ||:                  |||||||||||||| |||||||||| |||
Db    891  GA--------GGAGLNGNGGAGGNGGLGGFGGNGGAAGANGVAVGAPGQPGGAGGHGG 940

Qy    273  TFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADG 313
           :||
Db    941  AGGNGGAGGNGGQG--------VVSDGAGGAGGAGGDGGAPG 974

RESULT  10
E70806
hypothetical glycine-rich protein Rv3507 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70806
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Skelton, K.; Seeger, K.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A;Reference number: A70500; MUID:98295987
```

```
Db    192  AGGSGGNGGAGGWIYGHGGHGGAGGNGGNA-TAPGGASAGFDGGAGGNG------GSGGRG 245

Qy    123  GGQIGDNPLLKAMLKLIARMMDGQSDOFGQPGTGNNSASSGTSSSGGSPFND-------- 174
           |:                                          | | | |||
Db    246  GLLFGNG------------------GNGSVGGMGGQGTNDTAGDSAGSGLGGNGNGAQGGW 290

Qy    175  LSGGKAPSGNSPSGNYSPVSTFSPPSTPSTPLDFPSSPTKAAGGSTPVTDHPDPVGS 234
                 | || |  :| | || || |||||||| |||||||| :||||||
Db    291  LIGNGGQGGDSGAGG------------GTDSTQTGVMN-----GASGGSAGIAGNGGDAGL 334

Qy    235  AGIG-----AGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQ 289
               |                                             |
Db    335  VGNGGAGGNGGNGAAGSALGTT-------------------IFGGSSGVGGSGGDGGNGG----- 375

Qy    290  KPLFLEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVE 349
           ||| |||||||||||||||||||||||||||||||||||  ||| | |||||  | |
Db    376  ---WLFGSGAS-------GGNGGQG---GDA-------GTNGFAGFGGSAGGGGWVG 412

Qy    350  ITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGN-WDLNLSHISAEDG 407
             |                                          |||| : ::|:|
Db    413  AVN----------------------------FGPISVQGFGLF--GHGGDGGNGGDVGAGSLSIQFG 449

RESULT  8
B70523
hypothetical glycine-rich protein Rv0297 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
C;Accession: B70523
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Skelton, K.; Seeger, K.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: B70523
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-591 <COL>
A;Cross-references: GB:296800; GB:AL123456; NID:g3261800; PIDN:CAB09596.1; PID:e321655;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0297
C;Superfamily: unassigned collagens

Query Match       8.1%;   Score 186.5; DB 2;   Length 591;
Best Local Similarity   23.1%;   Pred. No. 0.0029;
Matches   91;  Conservative   31;  Mismatches 119;  Indels 153;  Gaps  15;

Qy      6  LNNNTSSSPGLFQSGGDNGL---GGHNANSALGGQQPIDRQTIEQMAQLLAELLKSLLSP 61
               | :||  ||||| || || ||                        ::||
Db    116  IGNGANGAPGTGQAGGDGGLLFGNGGNGGSSCAPGQ-----------TGAGGNDQTTGVNAG 152

Qy     62  QSGNAA----------------------TGAGGNDQTTGVNAG------------ 83
                                         ||| ||| ||
Db    153  GAGGAAGFFGNGGNGGDGGAGGANGGAGGTAGWFFGFGNGNGGAGGIGVAGINGGLGGAGGD 212

Qy     84  -GLNGRKGTAG------------TTPQDSQNMLSEMGNNGLDQAITPDGQGGQ 123
            |||:|:                 |||:|
Db    213  GGNAGFFGNGNGGMGGAGGAAGVNAVNPGLATPVTPAAN-----GGNGLNLVGPGTAGG 267

Qy    124  GGQIGDNPLLKAMLKLIARMMDGQSDOFGQPGTGNNSASSG-----TSSSGGSPFNDLSG 177
           ||| |||                             |:|||     :|||
Db    268  GADGAN----------GSAIGQAG-GAGDGGNASTSGGIGIAQTGGAGGA--GGAGG 312

Qy    178  GKAPSGNSPSGNYSPVSTFSPPSTPSTPLDFPSSPTKAAGGSTPVTDHPDPVGSAGI 237
                 | || |  :| | || || ||||||||||| :|||||| ||||
Db    313  DGAPGGNGGNG----------GSVEHTGATGSSASGNGATGGNGGVGAPG- 353

Qy    238  GAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILED 297
           |||                                             || | ||
```

A;Accession: E70806
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1381 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PID:CAA17744.1; PID:g292444
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3507
C;Superfamily: collagen alpha 1(IV) chain

Query Match   8.0%;   Score 185.5; DB 2;   Length 1381;
Best Local Similarity   25.0%;   Pred. NO. 0.0094;
Matches   79; Conservative   25; Mismatches 141; Indels   71; Gaps 11;

Qy   19 SGGDNGLGGHNANSALGQQ--PIDRQTIEQMAQLLAELLKSLLSPQSGNAATGGNDQT 76
Db   779 SGGDGGKGGQGGSSGCTGGSGAPI----------------GGGAGGTGGSGGH 814

Qy   77 TVGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGG-GQIGNDPLLKAM 135
Db   815 AGKGGAGGI-GAQGTTITVPGNGGN--AGDGGNGGNAGA---GGNGGSGDFGGNT----- 863

Qy   136 LKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKAPS-GNSPSGNYSPVS 194
Db   864 -----------TSGASGSGGNGGNAGTAGSGGAGTGGTGLSGGNGGNGGNGGDGGNG 913

Qy   195 TFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTV 254
Db   914 AHGTVGAQFVPATSLPTPNGGAGGNGGTGSNGGAPGPAGAPGPTTGGNAGSQGIGGDGGN 973

Qy   255 LHDTITVKAGQVFDG--------------KGQTFTAGS----ELGDGGQSENQKPLFILED 297
Db   974 GGD-----GGKGGDGADAVNVVFMPTEPQAATGTAGSAGDPTGGNGPGTPGSPMVAPPP 1028

Qy   298 GASLKNVTMGDDGADG 313
Db 1029 PTPITQVQQGGDGGAG 1044

RESULT 11
D70807
hypothetical glycine-rich protein Rv3514 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70807
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987
A;Accession: D70807
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1489 <COL>
A;Cross-references: GB:AL022022; GB:AL123456; NID:g3261554; PID:CAA17751.1; PID:g292445
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3514
C;Superfamily: collagen alpha 1(IV) chain

Query Match   8.0%;   Score 185.5; DB 2;   Length 1489;
Best Local Similarity   23.9%;   Pred. NO. 0.01;
Matches   85; Conservative   26; Mismatches 136; Indels 109; Gaps 12;

Qy   10 TSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATG 69
Db   726 TTGNAGVGGAGGSSGAGGTNGSGGAGG-------------TDGQGGAGGAG 763

Qy   70 AGGNDQTTGVG-----------NAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAIT 117

Db   764 GAGADNPTGIGTGGDGGDGGTGGEAAGAGGAGGAAGTGGT------GGMIGTTGNAGV----- 812
Qy   118 PDGQGGQIGNDPLLKAMLKLIARMMDGQSDQFGQGPG-TGNNSASSGTSSSGGSPFNDLS 176
Db   813 --GGAGGQGGDGG--------------AGGAGADADQPGATGGTGFAGGAGGAGGGGSSCA 858
Qy   177 GGKAPSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAG 236
Db   859 GGTNGSGGA-GGTCGGV-----------VAGGAGISFSNGSNGGTGGTGGVGTGGDGGNAG 908
Qy   237 IGA------------GNSVAFTSAGAN------QTVLHDTITVKAG 264
Db   909 TGAGDPGKGGTGGTGGTGGSGGAGGSGGANFNGGTGGTGGTGGKGGLNTDGLSSATSGTG 968
Qy   265 QV--FDGKGQTFTAGSEL-----GDGGQSENQKPLFILEDGASLKNVTMGDDGADG 313
Db   969 GTGGTGGKGGTGGAGDDSAGGTGGTGGAGGAGGLANTCGTAGNAGIGGDGGQG 1024

RESULT 12
A45724
pectate lyase (EC 4.2.2.2) - fungus (Fusarium solani)
C;Species: Fusarium solani
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
C;Accession: A45724
R:Gonzalez-Candelas, L.; Kolattukudy, P.E.
J. Bacteriol. 174, 6343-6349, 1992
A;Title: Isolation and analysis of a novel inducible pectate lyase gene from the phyt
A;Reference number: A45724; MUID:93015682
A;Accession: A45724
A;Status: preliminary
A;Molecule type: DNA; protein
A;Residues: 1-242 <GON>
A;Cross-references: GB:M94691; NID:g168155; PID:AAA33338.1; PID:g168156
A;Experimental source: isolate T8
A;Note: sequence extracted from NCBI backbone (NCBIN:115473, NCBIP:115474)
C;Keywords: carbon-oxygen lyase

Query Match   7.9%;   Score 183.5; DB 2;   Length 242;
Best Local Similarity   30.5%;   Pred. NO. 0.0014;
Matches   62; Conservative   32; Mismatches   80; Indels   29; Gaps   8;

Qy   235 AGIGAGNSVAFT------SAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGSENQ 289
Db   9 AALVGTSSAAVTKTLPKSAGA--TSFPTAVPVKGS--YDGMKRFEREPKYCKGQDETGE 64
Qy   290 K-PLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHV 348
Db   65 KDAMFLENGATLSNVIIGASQAEGVHCKGTCTLNNVWADVCEDAVTLKQTSGTS---- 120
Qy   349 EITNSSFEHASDKIIQLNADTNLSVDNVKAKDFGTFVRT------NGGQQGNWDLNLSHI 402
Db   121 YINGGGAFHASDKIIQFNGRGTVHVKDFYAEDYGKLSRSCGNCKDNGGPR--NVIVENS 177
Qy   403 SAEDGKFSFVKSDSEGLNVNTSD 425
Db   178 VAVDGGVLC------GINTNYGD 194

RESULT 13
F70971
hypothetical glycine-rich protein Rv3367 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 18-Feb-2000
C;Accession: F70971
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

```
A:Reference number: A70500; MUID:98295987
A:Accession: F70971
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-588 <COL>
A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15752.1; PID:e120228
A:Experimental source: strain H37Rv
A:Genetics:
A:Gene: Rv3367
C:Superfamily: unassigned collagens

Query Match      7.9%;    Score 183;   DB 2;   Length 588;
Best Local Similarity   23.1%;   Pred. No. 0.0046;
Matches   83;  Conservative  20;  Mismatches 151;  Indels 106;  Gaps  11;

Qy     7  NNNTSSSPGLFQSGGDNGLGGHHN------------------ANSALGQQPIDRQTIEQMAQL  50
              :  |||  ||    :||       ||
Db   208  NGGAGGAAGLWGSGSGGGGQGNGLTGNDGVNPAPVTNPALNGAAGDSNIEPQT--------  260

Qy    51  LAELLKSLLSPQSGNAATG----------------------AGGNDQTTGVGNA------  85
              :  |||                                ||
Db   261  -----SVLIGTQGGDTPGGACVNGGNGGAGGDANGNPANTSIANAGAGGNAAGGDGGA  315

Qy    86  NGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGG----GQIGDNPLLKAMLKLIAR  141
             || || ||    :: ||| |||    ||
Db   316  NGGAGGAGGQQAASAGSSVGGDGGNGGAGGTGTNGHAGGAGGAGGAGGR----------GG  365

Qy   142  MMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPST  201
            ||  ||  ||
Db   366  WLVGNGNGGNGAAGGNGAIGGTGGAGGVPAN--QGGNSALGTQPVGG----DGGDGGNG  419

Qy   202  PTSPTSPLDFPSSPTKAAG------GSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVL  255
             ||  ||||    :: |   :  |    | |
Db   420  GTGGTCGRGGDGGAGGASGWLMGNGNGNGGTGTGGSGGVGGNGIGGDGAGGGNATS  479

Qy   256  HDTITVKAGVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIH  315
              || ||  |            ||         ||     |||
Db   480  TSSIP------FDAHGNGGAGGDAGHGG----------------------TGGDGDGGH  512
```

```
RESULT  14
H70816
hypothetical glycine-rich protein Rv0872c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 21-Jan-2000
C:Accession: H70816
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: H70816
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-606 <COL>
A:Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17678.1; PID:e125401
A:Experimental source: strain H37Rv
A:Genetics:
A:Gene: Rv0872c
C:Superfamily: unassigned collagens

Query Match      7.9%;    Score 182;   DB 2;   Length 606;
Best Local Similarity   26.4%;   Pred. No. 0.0054;
Matches   90;  Conservative  27;  Mismatches 112;  Indels 112;  Gaps  15;

Qy    16  LFQSGGDNGLGGHHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGGNDQ  75
            |  ||  |   ||
Db   220  LFGSGAGGGQGMGAAGADGVNP----------------------TPTGTADAGSTGTDQ  257
```

```
Qy    76  TTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQG-GGQIGDNPLLKA  134
            |    ||  ||    |||    ||
Db   258  TLG-GNAIGGNGGPGDAGDAMTSGG----------AGGSGGNAVSTVNGDAVGGEGGKG---  305

Qy   135  MLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKAPS----GNSPSGN  189
                                ::
Db   306  --------GEGAYGGAGGAGGSAASIGNAAIGG---NGGAGGNAQAPGGVGGAGGEGG  352

Qy   190  YSPVSTFSPPSTPT---SPTSPLD-FPSSPTKAAGGS-----  222
              :  |||||  :: |   |||
Db   353  DAQVGTNSPSNAEAGNGGSGGNGFDSFASGGTGGAGGTGGAGGRGGLLIGDGGAGGAGGV  412

Qy   223  --TPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKG-------QT  273
             || | |           |||:::::
Db   413  GGTGGSGGAPGGGGGAG-GDGGAANTDSAGSSR--------KAFGGDGVGGDGASA  459

Qy   274  FTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGI  314
             | | |               |||
Db   460  LGTGGEGGIGGQGGN----------GGAG--GLLIGNGGAGGV  490
```

```
RESULT  15
C70974
hypothetical glycine-rich protein Rv3388 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: C70974
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: C70974
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-731 <COL>
A:Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15773.1; PID:g266
A:Experimental source: strain H37Rv
A:Genetics:
A:Gene: Rv3388
C:Superfamily: elastin

Query Match      7.9%;    Score 182;   DB 2;   Length 731;
Best Local Similarity   25.3%;   Pred. No. 0.0067;
Matches   89;  Conservative  26;  Mismatches 127;  Indels 110;  Gaps  15;

Qy     7  NNNTSSSPGLFQSGGDNGLGGHHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNA  66
             |  |  |                                    :
Db   399  NNTSAGTGGVGASGGTGGNAG----------------------LIGAGGHGG  428

Qy    67  ATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQ-----SDSQN----MLSEMGNNGLDQAIT  117
           | ||||  :  |||                       SDSQN
Db   429  AGGAGGN-QTGGVGN-GGAGGNGGAGGAGGQLYGNGGDGGNGGAGGANIAGGNGSDGGAA  486

Qy   118  PDQGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSAS------SGTSSSGGSSP  171
            |||   |                                   :  :
Db   487  GHGGAGGS----------ARLI-GAGGHGGDGGAGGNTAGRRADAIAGTGDGGNG  531

Qy   172  FNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDP  231
                                                              :
Db   532  GN----GGLLSGNAGAGHGGAGGSSTATTTTG---------TPPTGATGGN------  570

Qy   232  VGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKP  291
            | | ||                                     ||
Db   571  -GGNG-GAGGTAGFTGSG--------GIGGNGGAGGTGGNAGVA------  604

Qy   292  LFILEDGASLKNVTMGDDGADGITHLYGDAKIDNLHVTNVGEDAITVKPNSAG  343
             :|:
Db   605  LSVGSTGLGLGNGGSGGLGGGLFGNGGAGGVGATG-GNGGSGIGPASVG  655
```

us-09-596-958-2.rpr

Mon Oct 22 09:34:59 2001

THIS PAGE BLANK (USPTO)

```
              GenCore version 4.5
         Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:      October 18, 2001, 09:08:45 ; Search time 12.52 Seconds
                                           (without alignments)
                                           735.135 Million cell updates/sec

Title:            US-09-596-958-2
Perfect score:    2310
Sequence:         1 MSILTLNNTSSSPGLFQSG.........LGDVENHYKVPMSANLKVAE 447

Scoring table:    BLOSUM62
                  Gapop 10.0 , Gapext 0.5

Searched:         197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters:      197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :        Issued_Patents_AA:*
                  1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
                  2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
                  3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
                  4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
                  5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
                  6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

         Pred. No. is the number of results predicted by chance to have a
         score greater than or equal to the score of the result being printed,
         and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	559	24.2	424	4	US-09-120-817-2	Sequence 2, Appli
2	359.5	15.6	197	4	US-09-402-668-2	Sequence 2, Appli
3	341	14.8	221	4	US-09-198-956-4	Sequence 4, Appli
4	190.5	8.2	62	1	US-09-402-668-10	Sequence 10, Appli
5	165.5	7.2	344	1	US-08-891-254-7	Sequence 7, Appli
6	165.5	7.2	344	2	US-08-819-539-7	Sequence 7, Appli
7	165.5	7.2	344	4	US-09-030-270A-7	Sequence 7, Appli
8	165.5	7.2	344	4	US-08-984-207-7	Sequence 7, Appli
9	165.5	7.2	344	5	PCT-US96-08819-7	Sequence 7, Appli
10	160.5	6.9	907	5	US-09-010-928B-4	Sequence 4, Appli
11	156.5	6.8	1912	1	US-08-409-995-4	Sequence 4, Appli
12	156.5	6.8	1912	3	US-08-685-467-4	Sequence 4, Appli
13	155.5	6.7	2353	4	US-09-377-155-33	Sequence 33, Appli
14	155.5	6.7	2353	4	US-08-913-942-4	Sequence 4, Appli
15	155	6.7	385	5	PCT-US93-06243-2	Sequence 2, Appli
16	155	6.7	495	2	US-08-794-795-2	Sequence 2, Appli
17	155	6.7	495	4	US-09-249-200-2	Sequence 2, Appli
18	151	6.5	385	1	US-08-891-254-3	Sequence 3, Appli
19	151	6.5	385	2	US-08-819-539-3	Sequence 3, Appli
20	151	6.5	385	5	PCT-US96-08819-3	Sequence 3, Appli
21	151	6.5	403	2	US-08-200-724A-2	Sequence 2, Appli
22	151	6.5	403	4	US-09-030-270A-3	Sequence 3, Appli
23	151	6.5	403	2	US-08-851-376A-2	Sequence 2, Appli
24	151	6.5	403	4	US-08-984-207-3	Sequence 3, Appli
25	150.5	6.5	745	2	US-09-010-928B-28	Sequence 28, Appli
26	150.5	6.5	870	2	US-09-010-928B-2	Sequence 2, Appli
27	150	6.5	520	2	US-08-794-795-6	Sequence 6, Appli
28	150	6.5	520	4	US-09-249-200-6	Sequence 6, Appli
29	147.5	6.4	489	2	US-08-794-795-7	Sequence 7, Appli
30	147.5	6.4	489	4	US-09-249-200-7	Sequence 7, Appli
31	147.5	6.4	518	1	US-08-392-367B-2	Sequence 2, Appli
32	147.5	6.3	518	3	US-08-893-467A-2	Sequence 2, Appli
33	145	6.3	334	3	US-09-060-756-728	Sequence 728, App
34	144	6.2	1581	4	US-09-110-517-2	Sequence 2, Appli
35	143.5	6.2	679	4	US-08-913-942-15	Sequence 15, Appli
36	138	6.0	353	2	US-08-687-702-37	Sequence 37, Appli
37	137	5.9	300	3	US-08-765-856-2	Sequence 2, Appli
38	137	5.9	300	4	US-08-935-009A-2	Sequence 2, Appli
39	137	5.9	842	5	PCT-US96-02331-15	Sequence 15, Appli
40	136.5	5.9	1477	1	US-08-038-682-4	Sequence 4, Appli
41	136.5	5.9	1477	1	US-08-302-832-4	Sequence 4, Appli
42	136.5	5.9	1477	4	US-08-530-198-4	Sequence 4, Appli
43	136.5	5.9	1477	2	US-08-469-880-4	Sequence 4, Appli
44	136.5	5.9	1477	2	US-08-728-470-4	Sequence 4, Appli
45	136.5	5.9	1477	2	US-08-617-697-4	Sequence 4, Appli

ALIGNMENTS

```
RESULT   1
US-09-120-817-2
; Sequence 2, Application US/09120817
; Patent No. 6172184
; GENERAL INFORMATION:
;   APPLICANT:  Collmer, Alan
;   APPLICANT:  Charkowski, Amy
;   APPLICANT:  Alfano, James R.
;   TITLE OF INVENTION:  HYPERSENSITIVE RESPONSE ELICITOR FROM
;   TITLE OF INVENTION:  PSEUDOMONAS SYRINGAE AND ITS USE
;   NUMBER OF SEQUENCES:  8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;     STREET:  P.O. Box 1051, Clinton Square
;     CITY:  Rochester
;     STATE:  New York
;     COUNTRY:  U.S.A.
;     ZIP:  14603
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/120,817
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/055,107
;     FILING DATE:  06-AUG-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Goldman, Michael L.
;     REGISTRATION NUMBER:  30,727
;     REFERENCE/DOCKET NUMBER:  19603/1741
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (716) 263-1304
;     TELEFAX:  (716) 263-1600
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  424 amino acids
;     TYPE:  amino acid
;     STRANDEDNESS:
;     TOPOLOGY: linear
;   MOLECULE TYPE:  protein
US-09-120-817-2

  Query Match       24.2%;   Score 559;  DB 4;  Length 424;
  Best Local Similarity  36.1%;  Pred. No. 1.1e-36;
  Matches  147; Conservative  54; Mismatches  132; Indels  74; Gaps  13;
```

```
Qy    37  QPIDRQTIEQMAQLLAELLKSL---LSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAG 93
          ||:|::||  ::|:: |:  |:  :|  :|||
Db    72  KPNDSQS--NIAKLISALIMSLLQQMLTNSNKKQDTNQEQPDSQAPFQNNGGLG------ 122

Qy    94  TTPQDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMDGQSDQFGQP 153
          ||:||         |||| ||||:||||  |  ||:||
Db    123 -TPSADS------GGGG----TPDATGGGG-GDTP------------SATGGG 151

Qy    154 GTGNNSASSTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPS 213
          |||   :|: :||||  || : |:
Db    152 GGDTPTATGGGGSGSGGGGTPTATGGG---SGGTPTATGGGEGGVTPQITPQL-------A 200

Qy    214 SPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQT 273
          :|  |: || |:| |  ||||| || |:||
Db    201 NPNRTSG---------TGSVSDTAGS----TEQAGKINVVKDTIKVGAGEVFDGHGAT 245

Qy    274 FTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYG---DAKIDNLHVTN 329
          |||  :  ||   :||||||| ||: |  ||:| :||:  ::| |:  :|||:|:
Db    246 FTADKSMGNDGQGENQKPMFELAEGATLKNVNLGENEVDGIHVKAKNAQEVTIDNVHAQN 305

Qy    330 VGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNG 389
          ||| ||:|| :|  || || :|  :| | ||  | |||:: ::|| |||| |||| |||
Db    306 VGEDLITVKGEGGAAVTNLNIKNSSAKGADDKVVQLNANTHLKIDNFKADDFGTMVRTNG 365

Qy    390 GQQ-GNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLGDVENHY 435
          ||| ||:|||| |:|||||:||||| :|||:| |:||:: :|||  |
Db    366 GKQFDDMSIELNGIEANHGKFALVKSDSDDLKLATGNIAMTDVKHAY 412
```

RESULT 2
US-09-402-668-2
; Sequence 2, Application US/09402668
; Patent No. 6172030
; GENERAL INFORMATION:
; APPLICANT: WADA, Yasunao
; APPLICANT: KASAI, Miyuki
; APPLICANT: SHIKATA, Shitsuw
; APPLICANT: SUZUMATSU, Atsushi
; APPLICANT: KOIKE, Kenzo
; APPLICANT: HATADA, Yuji
; APPLICANT: KOBAYASHI, Tohru
; APPLICANT: ITO, Susumu
; APPLICANT: TSUMADORI, Masaki
; TITLE OF INVENTION: Detergent Composition
; FILE REFERENCE: 2173-0116P
; CURRENT APPLICATION NUMBER: US/09/402,668
; CURRENT FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 9-091142 JAPAN
; PRIOR FILING DATE: 1997-04-09
; PRIOR APPLICATION NUMBER: 9-242736 JAPAN
; PRIOR FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/01613
; PRIOR FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 197
; TYPE: PRT
; ORGANISM: Bacillus sp.
; OTHER INFORMATION: Strain: KSM-P15
US-09-402-668-2

Query Match 15.6%; Score 359.5; DB 4; Length 197;
Best Local Similarity 46.3%; Pred. No. 2.5e-21;
Matches 81; Conservative 29; Mismatches 56; Indels 9; Gaps 4;

```
Qy    253 TVLHDTITVKAGQVFDGKGQTFTAG-SELGDGGSENQKPLFILEDGASLKNVTMGDDGA 311
          |:||||||||||| ||||||||||   ||||||:||||||||||:|| |||||:|  |
Db    3   TVVHETIRVPAGQTFDGKGQTYVANPNTLGDGSQAENQKPIFRLEAGASLKNVVIGAPAA 62

Qy    312 DGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNL 371
```

```
Db    63  DGVHCYGDCTITNVIWEDVGEDALTLK--SSGT---VNISSGAAYKAYDKVFQINAAGTI 117

Qy    372 SVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSE---GLNVNT 423
          ::|:||:|  |  ||| |||   |:||||||| ||::|:||:|
Db    118 NIRNFRADDIGKLVRQNGGTTYKVVMVENCNISRVKDAILRTDSSTSTGRIVNT 172
```

RESULT 3
US-09-198-956-4
; Sequence 4, Application US/0198956
; Patent No. 6165769
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-198-956-4

Query Match 14.8%; Score 341; DB 4; Length 221;
Best Local Similarity 41.7%; Pred. No. 8.6e-20;
Matches 73; Conservative 33; Mismatches 59; Indels 10; Gaps 2;

```
Qy    254 VLHDTITVKAGQVFDGKGQTFTAGSELGDGGSENQKPLFILEDGASLKNVTMGDDGADG 313
          |:||:|| ||| ||||||:|:|  :|||||||||:||||||:||||:|||:||||||||
Db    31  VVHKTIVVEKGQTYDGKGKRLIAGPELGDGSOREDQKPIFKVEDGATLKNVLGAPAADG 90

Qy    314 IHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSV 373
          :|:||:|:: :|||||:|||   ||||:|||:|::
Db    91  VHTYGNASINNVVWEDVGEDALTVK-----SEGSVTINGGSARLAADKIFQINKASTFTV 145

Qy    374 DNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISL 428
          |  ||| ||| ||:|:|||   ||:| |: :|||||:| |:   :||:
Db    146 KNFTADQGGKFIRQLGGSTFKAVVNIDNCTITNMKEAIFRTDS-----STSSVTM 195
```

RESULT 4
US-09-402-668-10
; Sequence 10, Application US/09402668
; Patent No. 6172030
; GENERAL INFORMATION:
; APPLICANT: WADA, Yasunao
; APPLICANT: KASAI, Miyuki
; APPLICANT: SHIKATA, Shitsuw
; APPLICANT: SUZUMATSU, Atsushi
; APPLICANT: KOIKE, Kenzo
; APPLICANT: HATADA, Yuji
; APPLICANT: KOBAYASHI, Tohru
; APPLICANT: ITO, Susumu
; APPLICANT: TSUMADORI, Masaki
; TITLE OF INVENTION: Detergent Composition
; FILE REFERENCE: 2173-0116P
; CURRENT APPLICATION NUMBER: US/09/402,668
; CURRENT FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 9-091142 JAPAN
; PRIOR FILING DATE: 1997-04-09

```
; PRIOR APPLICATION NUMBER: 9-242736 JAPAN
; PRIOR APPLICATION FILING DATE: 1997-09-08
; PRIOR APPLICATION NUMBER: PCT/US98/01613
; PRIOR APPLICATION FILING DATE: 1998-04-09
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence
; OTHER INFORMATION: primer
US-09-402-668-10

Query Match             8.2%;   Score 190.5;  DB 4;  Length 62;
Best Local Similarity  65.0%;   Pred. No. 1.3e-08;
Matches    39;   Conservative    7;   Mismatches    13;   Indels    1;   Gaps    1;

Qy  253  TVLHDTITVKAGQVFDGKGQTFTAG-SELGDGGQSENQKPLFILEDGASLKNVTMGDDGA  311
             |:::|| ||| ||||||||| |::  : |||:|||||:||||||||||| :|  |
Db    3  TVVHETIRVPAGQTFDGKGQTYVANPNTLGDGSQAENQKPIFRLEAGASLKNVVIGAPAA   62

RESULT 5
US-08-891-254-7
; Sequence 7, Application US/08891254
; Patent No. 5776889
; GENERAL INFORMATION:
;   APPLICANT: Wel, Zhong-Min
;   APPLICANT: Beer, Steven V.
;   TITLE OF INVENTION: Hypersensitive Response
;   TITLE OF INVENTION: Induced Resistance In Plants
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle
;   STREET: Clinton Square, P.O. Box 1051
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 14603
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent In Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/891,254
;   FILING DATE: 10-JUL-1997
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/475,775
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Goldman, Michael L.
;   REGISTRATION NUMBER: 30,727
;   REFERENCE/DOCKET NUMBER: 14603/10050
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716) 263-1304
;   TELEFAX:   (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 344 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-891-254-7

Query Match             7.2%;   Score 165.5;  DB 1;  Length 344;
Best Local Similarity  32.0%;   Pred. No. 1.3e-05;
```

```
Matches    57;   Conservative    20;   Mismatches    80;   Indels    21;   Gaps    7;

Qy   18  QSGGD--NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN   73
             |::|  ||||||||| ||:|||||:|||||||||||  :|:|:||||:||||  ||||
Db  134  QPGGNDKGNCVGGANGAKAGGQGGLAEALQEIEQILAQLGGGAGAGGGVCGAGGA      193

Qy   74  DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGGQIGDN  129
            |:|:|| ||||||||| |     ||  ||  || |||| ||  :| ||     ||  ||
Db  194  DGGSGAGGAGGANGADGGNGVNGNQANGPQNAGDVNGANGAD-----DGSEDQGGLTGVL  248

Qy  130  PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP-FNDLSGGK     179
            :|  :|  : |:|  ||    : |:|   ||
Db  249  QKLMKILNALVQMMQQGGLGGGNQAGGSKGAGNASPASGANPGANQPGSADDQSGQ     306
```

```
RESULT 6
US-08-819-539-7
; Sequence 7, Application US/08819539
; Patent No. 5859324
; GENERAL INFORMATION:
;   APPLICANT: Wel, Zhong-Min
;   APPLICANT: Beer, Steven V.
;   TITLE OF INVENTION: Hypersensitive Response
;   TITLE OF INVENTION: Induced Resistance In Plants
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle
;   STREET: Clinton Square, P.O. Box 1051
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 14603
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patent In Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/819,539
;   FILING DATE: 17-MAR-1997
;   CLASSIFICATION: 800
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/475,775
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Goldman, Michael L.
;   REGISTRATION NUMBER: 30,727
;   REFERENCE/DOCKET NUMBER: 14603/10050
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716) 263-1304
;   TELEFAX:   (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 344 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-819-539-7

Query Match             7.2%;   Score 165.5;  DB 2;  Length 344;
Best Local Similarity  32.0%;   Pred. No. 1.3e-05;
Matches    57;   Conservative    20;   Mismatches    80;   Indels    21;   Gaps    7;

Qy   18  QSGGD--NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN   73
             |::|  ||||||||| ||:|||||:|||||||||||  :|:|:||||:||||  ||||
Db  134  QPGGNDKGNCVGGANGAKAGGQGGLAEALQEIEQILAQLGGGAGAGGGVCGAGGA      193

Qy   74  DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGGQIGDN  129
            |:|:|| ||||||||| |     ||  ||  || |||| ||  :| ||     ||  ||
Db  194  DGGSGAGGAGGANGADGGNGVNGNQANGPQNAGDVNGANGAD-----DGSEDQGGLTGVL  248
```

Qy 130 PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP--FNDLSGGK 179
           | : |: ||| : || | || || :   : ||| : || |:
Db 249 QKLMKILNALVQMMQQGGLGGGNQAQGGSKGAGNASPASGANPGANQPGSADDQSSGQ 306

RESULT 7
US-09-030-270A-7
; Sequence 7, Application US/09030270A
; Patent No. 5977060
; GENERAL INFORMATION:
;   APPLICANT: Zitter, Thomas A.
;   APPLICANT: Wei, Zhong-Min
;   TITLE OF INVENTION: INSECT CONTROL WITH A
;   TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR
;   NUMBER OF SEQUENCES: 10
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;   STREET: P.O. Box 1051, Clinton Square
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 14603
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/030,270A
;   FILING DATE:
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 60/039,226
;   FILING DATE: 28-FEB-1997
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Goldman, Michael L.
;   REGISTRATION NUMBER: 30,727
;   REFERENCE/DOCKET NUMBER: 19603/1521
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716) 263-1304
;   TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 344 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-030-270A-7

Query Match   7.2%; Score 165.5; DB 2; Length 344;
Best Local Similarity   32.0%; Pred. No. 1.3e-05;
Matches 57; Conservative 20; Mismatches 80; Indels `21; Gaps   7;

Qy   18 QSGGD---NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN 73
           | ||: ||:|||||||:||| || |:||::|||:|:| |||:|||| |:
Db 134 QPGGNDKGNGVGGANGAKGAGGQGGLAEALQEIEQILAQLGGGGAGGAGGVGGAGGA 193

Qy   74 DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGGQIGDN 129
          : :| ||| ||||| :||:| ||:|:|||||||||||:|||:|| || || ||
Db 194 DGGSGAGGAGGANGADGGNGVNGQNANGPQNAGDVNGANGAD----DGSEDQGGLTGVL 248

Qy 130 PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP--FNDLSGGK 179
           | : |: ||| : || | || || :   : ||| : || |:
Db 249 QKLMKILNALVQMMQQGGLGGGNQAQGGSKGAGNASPASGANPGANQPGSADDQSSGQ 306

RESULT 8
US-08-984-207-7
; Sequence 7, Application US/08984207
; Patent No. 6235974
;; GENERAL INFORMATION:

APPLICANT: Qiu, Dewen
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
TITLE OF INVENTION: RESISTANCE IN PLANTS BY SEED TREATMENT
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
  ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
  STREET: P.O. Box 1051, Clinton Square
  CITY: Rochester
  STATE: New York
  COUNTRY: U.S.A.
  ZIP: 14603
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/984,207
  FILING DATE:
  CLASSIFICATION:
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 60/033,230
  FILING DATE: 05-DEC-1996
ATTORNEY/AGENT INFORMATION:
  NAME: Goldman, Michael L.
  REGISTRATION NUMBER: 30,727
  REFERENCE/DOCKET NUMBER: 19603/1201
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (716) 263-1304
  TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
  LENGTH: 344 amino acids
  TYPE: amino acid
  STRANDEDNESS:
  TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-984-207-7

Query Match   7.2%; Score 165.5; DB 4; Length 344;
Best Local Similarity   32.0%; Pred. No. 1.3e-05;
Matches 57; Conservative 20; Mismatches 80; Indels 21; Gaps   7;

Qy   18 QSGGD---NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN 73
           | ||: ||:|||||||:||| || |:||::|||:|:| |||:|||| |:
Db 134 QPGGNDKGNGVGGANGAKGAGGQGGLAEALQEIEQILAQLGGGGAGGAGGVGGAGGA 193

Qy   74 DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGGQIGDN 129
          : :| ||| ||||| :||:| ||:|:|||||||||||:|||:|| || || ||
Db 194 DGGSGAGGAGGANGADGGNGVNGQNANGPQNAGDVNGANGAD----DGSEDQGGLTCVL 248

Qy 130 PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP--FNDLSGGK 179
           | : |: ||| : || | || || :   : ||| : || |:
Db 249 QKLMKILNALVQMMQQGGLGGGNQAQGGSKGAGNASPASGANPGANQPGSADDQSSGQ 306

RESULT 9
PCT-US96-08819-7
; Sequence 7, Application PCT/US9608819
; GENERAL INFORMATION:
;   APPLICANT: Cornell Research Foundation, Inc.
;   TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
;   TITLE OF INVENTION: RESISTANCE IN PLANTS
;   NUMBER OF SEQUENCES: 9
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;   STREET: Clinton Square, P.O. Box 1051
;   CITY: Rochester
;   STATE: New York
;   COUNTRY: U.S.A.

```
;   ZIP: 14603
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US96/08819
;   FILING DATE:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/475,775
;   FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
;   NAME: Goldman, Michael L.
;   REGISTRATION NUMBER: 30,727
;   REFERENCE/DOCKET NUMBER: 19603/10051
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (716) 263-1304
;   TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 344 amino acids
;   TYPE: amino acid
;   STRANDEDNESS:
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
PCT-US96-08819-7

Query Match        7.2%;    Score 165.5; DB 5;  Length 344;
Best Local Similarity  32.0%;   Pred. No. 1.3e-05;
Matches   57;   Conservative  20;  Mismatches  80;  Indels  21;  Gaps   7;

Qy    18 QSGGD--NGLGGHNANSALGQQPIDRQTIEQMAQLLAEL-LKSLLSPQSGNAATGAGGN  73
            |:|||  :|||||::|||||:::||| ||||:| :|  |||:|| :|
Db   134 QPGGNDKGNGVGGANGAKGAGGQGGLAEALQEIEQILAQLGGGGAGGAGGVGGAGGA 193

Qy    74 DQTTGVGNAGGLNGRKGTAGT--TPQSDSQNMLSEMGNNGLDQAITPDG--QGGGQIGDN 129
            || |||||:|| :|||      ||:|:|| :|||  | ||       |:|
Db   194 DGGSGAGGAGGANGADGGNGVNGQNAGDVNGANGAD----DGSEDQGGLTGVL 248

Qy   130 PLLKAMLKLIARMMD------GQSDQFGQPGTGNNSASSGTSSSGGSP-FNDLSGGK 179
            |  :||:||  ::||  :|     :|||:|   |||:::||  :||  :|:||
Db   249 QKLMKILNALVQMMQQQGLGGGNQAQGSKGAGNASPASGANPGGANQPGSADDQSSGQ 306
```

```
; ATTORNEY/AGENT INFORMATION:
;   NAME: Murphy Jr., Gerald M
;   REGISTRATION NUMBER: 28977
;   REFERENCE/DOCKET NUMBER: 1447-109P
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 907 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-09-010-928B-4

Query Match        6.9%;    Score 160.5; DB 2;  Length 907;
Best Local Similarity  21.4%;   Pred. No. 0.00012;
Matches   85;   Conservative  35;  Mismatches  134;  Indels  143;  Gaps  14;

Qy    12 SSPGLFQSGDGNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNA--ATG  69
            :||  :|| :|:|:||                                          -----AGGAGGSGG
Db   603 SGPGGYGPGG-SGAGGTGPGGAGG-----

Qy    70 AGGNDQTTGVGNAGGLNGRKGTAGTTPQSD--------------SQNMLSEMGNNGLDQ 114
            |||| :| :|||| :| ||| ::|             : |:  |:|
Db   635 AGGSGGAGGSGGAGGSGGVGGSGGTTTEDLDITIDGADGPITISEELTISGAGGSGPGG 694

Qy   115 A----ITPDGQGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGG- 169
            |              : :|   ||            :||   -------VSGPGG-VGPGGSGPGGVGSGGSGPGGV 740
Db   695 AGPGGVGPGGSGPGGVGPG-------

Qy   170 ----SPFNDLSGGKAPSGNSPSGN---------------YSPVSTFSP----------PST 201
            ----   |:  | :| :|  ||   :|: |:|
Db   741 GPGGYGPGGSGSGGVGPGGYGPGGSGGGFYGPGGSEPGYGPGSGTYGSGGYGPGAGGPYG 800

Qy   202 PTSPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITV 261
            | ||:| ::  :|  : |::  ||   : || |:|||::::|:||
Db   801 PGSPGGAYG-PGSPGGAYYPSSRVPDMVNGIMSAMQGSGFNY---------------- 841

Qy   262 KAGQVFDGKGQTFTAGSELGDGGQSENQRPLFILEDGASLKNVTMGDDGADGIHLYGDAK 321
            |                                       --------GTCNPNNVNVLMDAL 870
Db   842 --QMFGNMLSQYSSGS-------------

Qy   322 IDNLH-VTNVGEDAITVKPNSAGKKSHVEITNSSFEH 357
            : :|| :::|
Db   871 LAALHCLSNHGSSSFAPSPTPAAMSAYSNSVGRMFAY 907

RESULT 11
US-08-409-995-4
; Sequence 4, Application US/08409995
; Patent No. 5546259
; GENERAL INFORMATION:
;   APPLICANT: Barenkamp, Stephen I.
;   APPLICANT: St. Geme III, Joseph W.
;   TITLE OF INVENTION: Haemophilus Adhesion Proteins
;   NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;   STREET: Four Embarcadero Center, Suite 3400
;   CITY: San Francisco
;   STATE: CA
;   COUNTRY: USA
;   ZIP: 94111-4187
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/409,995
;   FILING DATE: 24-MAR-1995
;   CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Silva, Robin M. .
```

```
RESULT 10
US-09-010-928B-4
; Sequence 4, Application US/09010928B
; Patent No. 5994099
; GENERAL INFORMATION:
;   APPLICANT: Lewis, Randolph V
;   APPLICANT: Hayashi, Cheryl Y
;   TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
;   TITLE OF INVENTION: CODING THEREFOR
;   NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;   STREET: 8110 GATEHOUSE RD. SUITE 500E
;   CITY: FALLS CHURCH
;   STATE: VIRGINIA
;   COUNTRY: UNITED STATES OF AMERICA
;   ZIP: 22042
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/010,928B
;   FILING DATE: 22-JAN-1998
;   CLASSIFICATION: 435
```

```
              REGISTRATION NUMBER:  38,304
              REFERENCE/DOCKET NUMBER: A-61053/RFT
        TELECOMMUNICATION INFORMATION:
              TELEPHONE:  (415) 781-1989
              TELEFAX:  (415) 398-3249
              TELEX:  910 277299
INFORMATION FOR SEQ ID NO: 4:
        SEQUENCE CHARACTERISTICS:
              LENGTH:  1912 amino acids
              TYPE:  amino acid
              STRANDEDNESS:  double
              TOPOLOGY:  unknown
US-08-409-995-4
```

```
Query Match      6.8%;   Score 156.5; DB 1; Length 1912;
Best Local Similarity   19.7%;  Pred. No. 0.00068;
Matches 103;  Conservative 61; Mismatches 168; Indels 191; Gaps 22;

Qy   70 AGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQ--------NMLSEMGNNGLDQAITPDG 120
          :::| ||:  ||:  :|  ::|:  |             ::| || |||
Db 1209 SAGNKEITNVKSA--LKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDN 1266

Qy  121 QG-------GGQIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNN---SASSGTSSS 167
          ||       :||
Db 1267 NGKHTVTIDVAEAKVGDG---------LEKDTDGKIKLKVDNTDGNNLLTVDATKGASVA 1317

Qy  168 GGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPV-T 226
             :  |:|                       :  |:  |:|  |:|
Db 1318 ------------------KGEFNAVTT---DATTAQGTNANERGKVVVKGSNGATATET 1355

Qy  227 DHPD--PVGSAGIGAGNSVAF----------TSAGANQTV-LHDTITVKAGQVF- 267
          |||      :|                    ||  | ||:   || :|||:
Db 1356 DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAXDTLTLKAGKNLK 1415

Qy  268 ---DGKGQTFTAGSE-------------LGDGGQSENQKPLFILEDGASL-----KNVTMGD 308
            ||                        |:|
Db 1416 VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVN-----ITSDTKGLNFAKDSKTGD 1470

Qy  309 DGADGIHLYG-----------DAKIDNLHVTNVGEDAIT-----------VKPN 340
          |:  ||                :|||:|||:|||:|||                |||
Db 1471 DA--NIHLNGIASTLTDTLLNSGATTNLGGNGITDNEKKXAASVKDVLNAGWNVRGVKPA 1528

Qy  341 SAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFG----------- 382
          |||  :: || :|| ::| :|:|:  ||| |:|  :|: |:
Db 1529 SANN----QVENIDFVATYDTVDFVSGDKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHN 1584

Qy  383 ----------TFVRTNGGQQGN------------WDLNLSHISAEDGKF 409
                    ||:|  |||||             ||:|:| ||:|:|
Db 1585 GKLFTGKELKDANNNGVTVTETDGKDEGNGLVTAKAVIDAVNKAGWRVKTTGANGQNDDF 1644

Qy  410 SFVKS-----------DSEGLNVNTSDISLGDVENHYKVPMSANLKV 445
          |:|:            |:  ||:|  :|:| || |:: ||:  :|:
Db 1645 ATVASGTNVTFADGNGTTAEVTKANDGSITVKYNVKVADGLKL 1687
```

```
RESULT 12
US-08-685-467-4
; Sequence 4, Application US/08685467
; Patent No. 6060059
        GENERAL INFORMATION:
              APPLICANT:  St. Geme III, Joseph W.
              APPLICANT:  Barenkamp, Stephen J.
              TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
              NUMBER OF SEQUENCES:  6
        CORRESPONDENCE ADDRESS:
              ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
              STREET:  Four Embarcadero Center, Suite 3400
              CITY:  San Francisco
              STATE:  California
              COUNTRY:  United States
              ZIP:  94111-4187
        COMPUTER READABLE FORM:
```

```
              MEDIUM TYPE:  Floppy disk
              COMPUTER:  IBM PC compatible
              OPERATING SYSTEM:  PC-DOS/MS-DOS
              SOFTWARE:  PatentIn Release #1.0, Version #1.30
        CURRENT APPLICATION DATA:
              APPLICATION NUMBER:  US/08/685,467
              FILING DATE:  22-JUL-1996
              CLASSIFICATION:  424
        PRIOR APPLICATION DATA:
              APPLICATION NUMBER:  US 08/409,995
              FILING DATE:  24-MAR-1995
        ATTORNEY/AGENT INFORMATION:
              NAME:  Silva, Robin M.
              REGISTRATION NUMBER:  38,304
              REFERENCE/DOCKET NUMBER: A-61053-2/RFT/RMS
        TELECOMMUNICATION INFORMATION:
              TELEPHONE:  (415) 781-1989
              TELEFAX:  (415) 398-3249
              TELEX:  910 277299
INFORMATION FOR SEQ ID NO: 4:
        SEQUENCE CHARACTERISTICS:
              LENGTH:  1912 amino acids
              TYPE:  amino acid
              STRANDEDNESS:  unknown
              TOPOLOGY:  unknown
              MOLECULE TYPE:  protein
US-08-685-467-4
```

```
Query Match      6.8%;   Score 156.5;  DB 3;  Length 1912;
Best Local Similarity   19.7%;  Pred. No. 0.00068;
Matches 103;  Conservative 61; Mismatches 168; Indels 191; Gaps 22;

Qy   70 AGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQ--------NMLSEMGNNGLDQAITPDG 120
          :::| ||:  ||:  :|  ::|:  |             ::| || |||
Db 1209 SAGNKEITNVKSA--LKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDN 1266

Qy  121 QG-------GGQIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNN---SASSGTSSS 167
          ||       :||
Db 1267 NGKHTVTIDVAEAKVGDG---------LEKDTDGKIKLKVDNTDGNNLLTVDATKGASVA 1317

Qy  168 GGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPV-T 226
             :  |:|                       :  |:  |:|  |:|
Db 1318 ------------------KGEFNAVTT---DATTAQGTNANERGKVVVKGSNGATATET 1355

Qy  227 DHPD--PVGSAGIGAGNSVAF----------TSAGANQTV-LHDTITVKAGQVF- 267
          |||      :|                    ||  | ||:   || :|||:
Db 1356 DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAXDTLTLKAGKNLK 1415

Qy  268 ---DGKGQTFTAGSE-------------LGDGGQSENQKPLFILEDGASL-----KNVTMGD 308
            ||                        |:|
Db 1416 VKRDGKNITFALANDLSVKSATVSDKLSLGTNGNKVN-----ITSDTKGLNFAKDSKTGD 1470

Qy  309 DGADGIHLYG-----------DAKIDNLHVTNVGEDAIT-----------VKPN 340
          |:  ||                :|||:|||:|||:|||                |||
Db 1471 DA--NIHLNGIASTLTDTLLNSGATTNLGGNGITDNEKKXAASVKDVLNAGWNVRGVKPA 1528

Qy  341 SAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFG----------- 382
          |||  :: || :|| ::| :|:|:  ||| |:|  :|: |:
Db 1529 SANN----QVENIDFVATYDTVDFVSGDKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHN 1584

Qy  383 ----------TFVRTNGGQQGN------------WDLNLSHISAEDGKF 409
                    ||:|  |||||             ||:|:| ||:|:|
Db 1585 GKLFTGKELKDANNNGVTVTETDGKDEGNGLVTAKAVIDAVNKAGWRVKTTGANGQNDDF 1644

Qy  410 SFVKS-----------DSEGLNVNTSDISLGDVENHYKVPMSANLKV 445
          |:|:            |:  ||:|  :|:| || |:: ||:  :|:
Db 1645 ATVASGTNVTFADGNGTTAEVTKANDGSITVKYNVKVADGLKL 1687
```

```
RESULT 13
US-09-377-155-33
; Sequence 33, Application US/09377155
```

; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
;   LENGTH: 2353
;   TYPE: PRT
;   ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match   6.7%; Score 155.5; DB 4; Length 2353;
Best Local Similarity 19.7%; Pred. No. 0.0011;
Matches 103; Conservative 61; Mismatches 168; Indels 191; Gaps 22;

```
Qy   70  AGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQ--------NMLSEMGNNGLDQAITPDG 120
             : ||  | :|  | :| | |   ::    |    ||  | :
Db 1210  SAGNKEITNVKSA--LKTYKDTQNTADETQDKEFHAAVKNANEVEFVGKNGATVSAKTDN 1267

Qy  121  QG--------GGQIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNN----SASSGTSSS 167
              :||           |||    :| ::|
Db 1268  NGKHTVTIDVAEAKVGDG--------LEKDTDGKIKLKVDNTDGNNLLTVDATKGASVA 1318

Qy  168  GGSPFNDLSGGKAPSGNSPSGNYSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPV-T 226
             :|: |:|  :| |: | :|:  :|:|
Db 1319  ------------KGEFNAVTT---DATTAQGTNANERGKVVVKGSNGATATET 1356

Qy  227  DHPD--PVGSAGIGAGNSVAF-------------TSAGANQTV-LHDTITVKAGQVF- 267
              ||           :|             |||  ||::|:::
Db 1357  DKKKVATVGDVAKAINDAATFVKVENDDSATIDDSPTDDGANDALKAGDTLTLKAGKNLK 1416

Qy  268  ---DGKGQTFTAGSE------------------LGDGGQSENQKPLFILEDGASL---KNVTMGD 308
             ||                         |  || :||
Db 1417  VKRDGKNITFALANDLSVKSATVSDKLSLGTNGKVN----ITSDTKGLNFAKDSKTGD 1471

Qy  309  DGADGIHLYG------DAKIDNLHVTNVGEDAIT-----------VKPN 340
            | |||            | ||  || | ||
Db 1472  DA--NIHLNGIASTLTDTLLNSGATTNLGGNGITDNEKKRAASVKDVLNAGWNVRGVKPA 1529

Qy  341  SAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFG---------- 382
             ||   :     |: :|:    ::| |:  |
Db 1530  SANN---QVENIDFVATYDTVDFVSGDKDTTSVTVESKDNGKRTEVKIGAKTSVIKDHN 1585

Qy  383  ------------TFVRTNGGQQGN------------WDLNLSHISAEDGKF 409
                          || ::|   ||
Db 1586  GKLFTGKELKDANNNGVTVTETDGKDEGNGLVTAKAVIDAVNKAGWRVKTTGANGQNDDF 1645

Qy  410  SFVKS-----DSEGLNVNTSDISLGDVENHYKVPMSANLKV 445
             ||        :  ||    |:|:||
Db 1646  ATVASGTNVTFADGNGTTAEVTKANDGSITVKYNVKVADGLKL 1688
```

RESULT 14
US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Gene, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,942
; FILING DATE: 29-DEC-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/409,995
; FILING DATE: 24-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/4031
; FILING DATE: 22-MAR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Vance, Dolly A.
; REGISTRATION NUMBER: 39,054
; REFERENCE/DOCKET NUMBER: A-61053-1/RFT/RMS/DAV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2353 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-913-942-4

```
Qy   410 SFVKS------DSEGLNVNTSDISLGDVENHYKVPMSANLKV 445
             :  :     ||  ::  :  :|: ::|: ||:
Db  1646 ATVASGTNVTFADGNGTTAEVTKANDGSITVKYNVKVADGLKL 1688
```

```
Qy   296 EDGASLKNVTMGDDGADGIHLYGD-AKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSS 354
             :    :  | | :|  : ::| :  :  :|: :|: ||: ||
Db   232 --GTGLDGSSLGGKGKLRGLSGPVDYQQLGNAVGTGIGMKAGIQALNDIG--THRHSSTRS 287
```

```
Qy   355 FEHASDKILQLNADTNLSVDNVKAKDFGTFV 385
             | :    :   |: |:  ::
Db   288 FVNKGDRAM------------AKEIGQFM 304
```

Search completed: October 18,. 2001, 09:09:09
Job time: 24 sec

RESULT 15
PCT-US93-06243-2
; Sequence 2, Application PC/TUS9306243
; GENERAL INFORMATION:
; APPLICANT: Zhong-Min Wel, David W. Bauer, Steven V.
; APPLICANT: Beer, Alan Collmer, Sheng-Yang He, and Ron J. Laby.
; TITLE OF INVENTION: Elicitor of the Hypersensitive Response in Plants
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Yahwak & Associates
; STREET: 25 Skytop Drive
; CITY: Trumbull
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06611
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06243
; FILING DATE: 19930630
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 907,935
; FILING DATE: 01-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: George M. Yahwak
; REGISTRATION NUMBER: 26,824
; REFERENCE/DOCKET NUMBER: CRF D-1172
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203)268-1951
; TELEFAX: (203)268-1951
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-06243-2
```

```
Query Match        6.7%;   Score 155;  DB 5;  Length 385;
Best Local Similarity  22.3%;  Pred. No. 0.0001;
Matches   87;  Conservative   36;  Mismatches 138;  Indels 130;  Gaps  15;
```

```
Qy   19 SGGDNG-LGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGGNDQTT 77
            :||:|| ||    |    : ||| ||| :|
Db   20 AGNNGLLGTSRQNAGLG-----------------GNSALGLGGGNQND 51
```

```
Qy   78 GVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLK 137
           ||    ||  : || | ||  || ||
Db   52 TVNQLAGL--------LTGMMMMSMMGGGGL------MGGGLGG----------- 82
```

```
Qy  138 LIARMMDGQSDQFGQPGTGNNSASSGTSSSG-GSPFNDLSGGKAPSGNSPSGNYSPVSTF 196
              | ||  : ||  :|  : || : |: ||: : :|
Db   83 ------------GLGNGLGGSGGLGEGLSNALNDMLGGSLNTLGSKGGN---NTT 122
```

```
Qy  197 SPPSTPTSPTSPLDFPSSPTKAAGGSTPVTDHPDPV------------------GSAGIGAGN 241
              | ::| || :| ||      :: :| |||
Db  123 STTNSPLDQALGINSTSQNDSTSGTDSTSDSSDPMQQLLKMFSEIMQSLFGDGDGTQG 182
```

```
Qy  242 SVAFTSAGANQTVLHDTITVKAGQVFD--------GKGQTFTAGSELGDGGQSENQKPLFIL 295
           | | | |    :|:  | : |: | | :    ||| |
Db  183 S----SSGGKQPTEGEQNAYKKG-VTDALSGLMGNGLSQLLGNGLGGGQGGNA------ 231
```

```
                    GenCore version 4.5
          Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on:          October 18, 2001, 09:08:46 ; Search time 24.08 Seconds
                                               (without alignments)
                                               2455.994 Million cell updates/sec

Title:           US-09-596-958-2
Perfect score:   2310
Sequence:        1 MSILTLNNNTSSSPGLFQSG.........LGDVENHYKVPMSANLKVAE 447

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters:    425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_16:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_unclassified:*
                13: sp_vertebrate:*
                14: sp_virus:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2310	100.0	447	2	O54508	O54508 erwinia amy
2	2277	98.6	447	2	Q9LAW2	Q91aw2 erwinia amy
3	567.5	24.6	441	2	O87327	O87327 pseudomonas
4	559	24.2	424	2	O87264	O87264 pseudomonas
5	362.5	15.7	224	2	Q9RHW0	Q9rhw0 bacillus sp
6	353.5	15.3	221	2	O34310	O34310 bacillus su
7	349.5	15.1	222	2	Q9X622	Q9x622 bacillus su
8	226.5	9.8	266	2	Q9EX16	Q9ex16 streptomyce
9	216.5	9.4	215	3	O00843	O00843 nectria hae
10	207	9.0	1660	2	O53215	O53215 mycobacteri
11	199.5	8.6	1306	2	O53775	O53775 mycobacteri
12	194.5	8.4	233	3	O00845	O00845 nectria hae
13	193	8.4	240	3	O93877	O93877 fusarium ox
14	193	8.4	694	2	O53212	O53212 mycobacteri
15	191	8.3	929	5	Q9NBW0	Q9nbw0 drosophila
16	191	8.3	929	5	Q9NBL3	Q9nbl3 drosophila
17	191	8.3	939	5	Q9NHQ0	Q9nhq0 drosophila
18	191	8.3	2280	5	Q9V8E6	Q9v8e6 drosophila
19	191	8.3	2302	5	Q9N693	Q9n693 drosophila

20	188	8.1	639	2	O53952	O53952	mycobacteri
21	186.5	8.1	591	2	O07224	O07224	mycobacteri
22	186	8.1	929	5	Q9NGW5	Q9NGW5	drosophila
23	186	8.1	1329	2	O06810	O06810	mycobacteri
24	185.5	8.0	809	5	P90534	P90534	dictyosteli
25	185.5	8.0	1381	2	O53552	O53552	mycobacteri
26	185.5	8.0	1489	2	O53559	O53559	mycobacteri
27	185	8.0	2310	5	Q9GRA9	Q9gra9	drosophila
28	183.5	7.9	242	3	Q04701	Q04701	fusarium so
29	183	7.9	588	2	O50396	O50396	mycobacteri
30	182.5	7.9	244	3	O00851	O00851	nectria hae
31	182	7.9	606	2	O53884	O53884	mycobacteri
32	182	7.9	731	2	O50415	O50415	mycobacteri
33	181.5	7.9	484	2	O53394	O53394	mycobacteri
34	180.5	7.8	1079	2	O53557	O53557	mycobacteri
35	179.5	7.8	588	14	Q9QEK6	Q9QEK6	cynomolgus
36	179	7.7	730	10	Q9ZU23	Q9ZU23	arabidopsis
37	179	7.7	1538	2	O53395	O53395	mycobacteri
38	179	7.7	3016	2	P73590	P73590	synechocyst
39	178	7.7	741	2	O06808	O06808	mycobacteri
40	177.5	7.7	2232	5	P91365	P91365	caenorhabdi
41	176.5	7.6	562	2	O50458	O50458	mycobacteri
42	176.5	7.6	2090	5	Q9W2T1	Q9W2T1	drosophila
43	175	7.6	645	2	O53818	O53818	mycobacteri
44	174	7.5	615	2	O05806	O05806	mycobacteri
45	174	7.5	2586	5	Q9VTK8	Q9VTK8	drosophila

```
                              ALIGNMENTS

RESULT    1
O54508
ID   O54508           PRELIMINARY;       PRT;     447 AA.
AC   O54508;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   HRPW PROTEIN.
GN   HRPW.
OS   Erwinia amylovora.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Erwinia.
OX   NCBI_TaxID=552;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CFBP1430;
RX   MEDLINE=98086111; PubMed=9426142;
RA   Gaudriault S., Malandrin L., Paulin J.P., Barny M.A.;
RT   "DspA, an essential pathogenicity factor of Erwinia amylovora showing
RT   homology with AvrE of Pseudomonas syringae, is secreted via the Hrp
RT   secretion pathway in a DspB-dependent way.";
RL   Mol. Microbiol. 26:1057-1069(1997).
RN   [2]
RP   SEQUENCE OF 138-447 FROM N.A.
RC   STRAIN=CFBP1430;
RX   MEDLINE=9806611; PubMed=9426142;
RA   Gaudriault S., Brisset M.N., Barny M.A.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 1-138 FROM N.A.
RC   STRAIN=EA321;
RX   MEDLINE=98115919; PubMed=9448330;
RA   Bogdanove A.J., Kim J.F., Wei Z., Kolchinsky P., Charkowski A.O.,
RA   Conlin A.K., Collmer A., Beer S.V.;
RT   "Homology and functional similarity of an hrp-linked pathogenicity
RT   locus, dspEF, of Erwinia amylovora and the avirulence locus avrE of
RT   Pseudomonas syringae pathovar tomato.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:1325-1330(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EA321, ATCC 49947;
RA   Kim J.F., Zumoff C.H., Beer S.V.;
RT   "HrpW, a new harpin of Erwinia amylovora, is a member of a family of
```

```
RT       pectate lyases.";
RL       Phytopathology 87:0-0(1997).
RN       [5]
RC       SEQUENCE FROM N.A.
RC       STRAIN=EA321, ATCC 49947;
RX       MEDLINE=98422475; PubMed=9748455;
RA       Kim J.F., Beer S.V.;
RT       "HrpW of Erwinia amylovora, a new harpin that contains a domain
RT       homologous to pectate lyases of a distinct class.";
RL       J. Bacteriol. 180:5203-5210(1998).
DR       EMBL; Y13831; CAA74158.1; -.
DR       EMBL; U97504; AAC04849.1; -.
DR       EMBL; U94513; AAC62314.1; -.
SQ       SEQUENCE   447 AA;  45347 MW;  7A744D169D6D9CF3 CRC64;

Query Match             100.0%;  Score 2310;  DB 2;  Length 447;
Best Local Similarity   100.0%;  Pred. No. 2.3e-137;
Matches  447;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSILTLNNTSSSPGLFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSILTLNNTSSSPGLFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60

Qy         61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG 120

Qy        121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180

Qy        181 PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240

Qy        241 NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300

Qy        301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360

Qy        361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420

Qy        421 VNTSDISLGDVENHYKVPMSANLKVAE 447
              |||||||||||||||||||||||||||
Db        421 VNTSDISLGDVENHYKVPMSANLKVAE 447

RESULT  2
Q9LAW2
ID       Q9LAW2       PRELIMINARY;    PRT;    447 AA.
AC       Q9LAW2;
DT       01-OCT-2000 (TrEMBLrel. 15, Created)
DT       01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT       01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE       HARPIN HRPW.
GN       HRPW.
OS       Erwinia amylovora.
OC       Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC       Erwinia.
OX       NCBI_TaxID=552;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN=EA246;
RA       Kim J.F., Laby R.J., Beer S.V.;
RT       "Comparison of the hrpN-flanking regions of two Erwinia amylovora
RT       strains with different host specificity.";
RL       Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
```

```
DR       EMBL; AF083620; AAF63402.1; -.
SQ       SEQUENCE   447 AA;  45340 MW;  0BBAEA3871EDC2F6 CRC64;

Query Match             98.6%;  Score 2277;  DB 2;  Length 447;
Best Local Similarity   98.7%;  Pred No. 2.7e-135;
Matches  441;  Conservative    3;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MSILTLNNTSSSPGLFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60
              ||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSILTLNSTSSSPGLFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLS 60

Qy         61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG 120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDG 120

Qy        121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QGGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKA 180

Qy        181 PSGNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db        181 PSGNSPSGSYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAG 240

Qy        241 NSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300
              |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db        241 NSVAFTSAGANQTVLHDTITVKASQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGAS 300

Qy        301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 LKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASD 360

Qy        361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLN 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 KILQLNADTNLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSHISAEDGKFSFVKSDSEGLK 420

Qy        421 VNTSDISLGDVENHYKVPMSANLKVAE 447
              |||||||||||||||||||||||||||
Db        421 VNTSDISLGDVENHYKVPMSANLKVAE 447

RESULT  3
O87327
ID       O87327       PRELIMINARY;    PRT;    441 AA.
AC       O87327;
DT       01-NOV-1998 (TrEMBLrel. 08, Created)
DT       01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT       01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE       HRP TYPE III SECRETED PROTEIN.
GN       HRPW.
OS       Pseudomonas syringae.
OC       Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC       Pseudomonas.
OX       NCBI_TaxID=317;
RN       [1]
RC       SEQUENCE FROM N.A.
RC       STRAIN=B728A;
RX       MEDLINE=98422476; PubMed=9748456;
RA       Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA       Collmer A.;
RT       "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT       to harpins and pectate lyases and can elicit the plant hypersensitive
RT       response and bind to pectate.";
RL       J. Bacteriol. 180:5211-5217(1998).
DR       EMBL; AF037983; AAC62530.1; -.
SQ       SEQUENCE   441 AA;  44714 MW;  7B3B3146E51897BB CRC64;

Query Match             24.6%;  Score 567.5;  DB 2;  Length 441;
Best Local Similarity   36.4%;  Pred. No. 3.5e-28;
Matches  156;  Conservative   46;  Mismatches  133;  Indels   93;  Gaps   14;

Qy         36 QQPIDRQTI----EQMAQLLAELLKSLLS--PQSGNAATGAGGNDQTTGVGNAGGLNGR 88
```

```
Db   67  QNPTDSSAATDPQSNVKLLSALVTSLLQMLMNLNKKQDTGQDSNEWQDPFQNEGGLG--  124
         ||      :  :||: |:||        |:|  ||              ||
Qy   89  KGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLLIARMMDGQSD  148
                  :|                                  :|
Db  125  -----TPSA---------------EGSDGGT---------------QEA            138

Qy  149  QFGQPGTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPS----GNYSPVST----FSPPS  200
              |  |||   ||  ||                         | |  ||      ||
Db  139  SGGDEGGGTTAATGDGGGGSTPTTEGDGG----GTSPTAEGDGGGSYVSTGADGSGAPS   194

Qy  201  TPPSPTSPLDFPSSPTKAAGGSTPVTDHPDP-VGSAGIGAGNSVAFTSAGANQ-----TV  254
         |               |                  |  |  :|  :    |
Db  195  T-------EDCTGGGGSDGVTPQVTPQVLANPGRNSGNGTVSDTTGSLEQSGEVNV      243

Qy  255  LHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGI  314
         |  ||  | |||||| | :||           :|||||  | |||  ||||| || |||
Db  244  VKDTIKVGAGQVFDGHGATFTADKSMGTGDQDEHQKPLFELAEGAVLKNVLGENEADGI   303

Qy  315  HLYGDAK------IDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNAD  368
         |          | ||| || ||||| ||||   | :| ||| |||| ||||| :|||||
Db  304  HV--NAKNSEQVTIDNVHAQNVGEDMITVKGEGGAKVTNLNITNSSANGADDKVIQLNAD  361

Qy  369  TNLSVDNVKAKDFGTFVRTNGGQQ-GNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDIS  427
         :  | || ||||||| ||||||:| |  |     || |||| || |||||||||||:|:|
Db  362  THLKVDGFKATDFGTLVRTNGGKQFDFDDMSVELNGVDATHGKFALVKSDSDDLKLATGDIA 421

Qy  428  LGDVENHY  435
         | |: |||
Db  422  MTDVKHAY  429
```

RESULT 4
O87264

```
ID   O87264      PRELIMINARY;      PRT;      424 AA.
AC   O87264;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   HRPW.
GN   HRPW.
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC   Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=98422476; PubMed=9748456;
RA   Charkowski A.O., Alfano J.R., Preston G., Yuan J., He S.Y.,
RA   Collmer A.;
RT   "The Pseudomonas syringae pv. tomato HrpW protein has domains similar
RT   to harpins and pectate lyases and can elicit the plant hypersensitive
RT   response and bind to pectate.";
RL   J. Bacteriol. 180:5211-5217(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=20243785; PubMed=10781092;
RA   Alfano J.R., Charkowski A.O., Deng W.L., Badel J.L.,
RA   Petnicki-Ocwieja T., van Dijk K., Collmer A.;
RT   "The Pseudomonas syringae Hrp pathogenicity island has a tripartite
RT   mosaic structure composed of a cluster of type III secretion genes
RT   bounded by exchangeable effector and conserved effector loci that
RT   contribute to parasitic fitness and pathogenicity in plants.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:4856-4861(2000).
DR   EMBL; AF005221; AAC62526.1; -.
DR   EMBL; AF232006; AAF71503.1; -.
SQ   SEQUENCE   424 AA;   42910 MW;   7AF4ED059BE79D2E CRC64;
```

Query Match   24.2%;
Best Local Similarity   36.1%;   Score 559;   DB 2;   Length 424;
Pred. No. 1.1e-27;

Matches 147;   Conservative 54;   Mismatches 132;   Indels 74;   Gaps 13;

```
Qy   37  QPIDRQTIEQMAQLLAELLKSL---LSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAG  93
         ||                            ||     ||            | |
Db   72  KPNDSQS--NIAKLISALIMSLLQMLTNSNKKQDTNQEQPDSQAPFQNNGGLG------  122

Qy   94  TTPQSDSQNMLSEMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLLIARMMDGQSDQFGQP  153
         || ||                  |||  |||     ||       || ||
Db  123  -TPSADS----GGGG---TPDATGGGG-GDTP-------------SATGGG          151

Qy  154  GTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPLDFPS    213
          |  ||      | ||     |||  | |                    ::|
Db  152  GGDTPTATGGGGSGGGGTPTATGGG---SGGTPTATGGGEEGVTPQITPQL------A   200

Qy  214  SPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQT  273
          |     ||   |           |           |        |  |  | |||| |
Db  201  NPNRTSG----TGSVSDTAGS----TEQAGKINVVKDTIKVGAGEVFDGHGAT        245

Qy  274  FTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYG----DAKIDNLHVTN  329
         |||    | |:|||||||:|  | ||| ||| |  | |||||| |       |||| | |
Db  246  FTADKSMGNGDQGENQKPMFELAEGATLKRVNLGENEVDGIHVKAKNAQEVTIDNVHAQN  305

Qy  330  VGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNG  389
         |||| ||||   |  || | ||||   |  || ||||| ::| || ||||||| |||||
Db  306  VGEDLITVKGEGGAAVTNLNIKNSSAKGADDKVVQLNANTHLKIDNFKADDFGTMVRTNG  365

Qy  390  GQQ-GNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLGDVENHY  435
         |:  |  | |     || |||| | |||| |::      | |: |||
Db  366  GKQFDDMSIELNGIEANHGKFALVKSDSDDLKLATGNIAMTDVKHAY  412
```

RESULT 5
Q9RHW0

```
ID   Q9RHW0      PRELIMINARY;      PRT;      224 AA.
AC   Q9RHW0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   PECTATE LYASE.
OS   Bacillus sp.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1409;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KSM-P15;
RA   Hatada Y., Koike K., Saito K., Kobayashi T., Susumu T.;
RT   "Amino acid sequence and possible catalytic residues of a novel
RT   alkaline pectate lyase from alkaliphilic Bacillus.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011839; BAA87892.1; -.
KW   Lyase.
SQ   SEQUENCE   224 AA;   23751 MW;   99D04821B09DE523 CRC64;
```

Query Match   15.7%;
Best Local Similarity   46.1%;   Score 362.5;   DB 2;   Length 224;
Pred. No. 1.1e-15;

Matches 83;   Conservative 29;   Mismatches 59;   Indels 9;   Gaps 4;

```
Qy  248  AGANQTVLHDTITVKAGQVFDGKGQTFTAG-SELGDGGQSENQKPLFILEDGASLKNVTM  306
         |     || |  | |:|| ||||||| :|     || |:|:|||| |   :|||||||:
Db   25  AEAAPTVVHETIRVPAGQTFDGKGQTYVANPNTLGDGSQAENQKPIFRLEAGASLKNVVI  84

Qy  307  GDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLN  366
          :  |||||  || ||  |  ::|||| :|||      | :|        |  |   ||
Db   85  GAPAADGVHCYGDCTITNVIWEDVGEDALTLK--SSGT---VNISGGAAYKAYDKVFQIN  139

Qy  367  ADTNLSVDNVKAKDFGTFVRTNGGQGWDLNLSHISAEDGKFSFVKSDSE---GLNVNT    423
         |  || |:| :| |||| ||:||||          |  |||| :|||    |:|||:
Db  140  AAGTINIRNFRADDIGKLVRQNGGTYKVMNVENCNISRVDAILRTDSSTSGRIVNT      199
```

RESULT 6

```
O34310
ID   O34310      PRELIMINARY;      PRT;   221 AA.
AC   O34310;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   YVPA.
GN   YVPA.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lazarevic V., Soldo B., Rivolta C., Reynolds S., Mauel C.,
RA   Karamata D.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RX   MEDLINE=98044033; PubMed=9384377;
RA   Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA   Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA   Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA   Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA   Chol S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA   Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA   Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA   Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA   Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA   Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA   Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA   Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA   Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,
RA   Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA   Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA   Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA   Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA   Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA   Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA   Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA   Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA   Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA   Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA   Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA   Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA   Viari A., Wambutt R., Wedler E., Wedler H., Weitzenegger T.,
RA   Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA   Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT   "The complete genome sequence of the gram-positive bacterium Bacillus
RT   subtilis.";
RL   Nature 390:249-256(1997).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=168;
RA   Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF017713; AAC67291.1; -.
DR   EMBL; Z99121; CAB15500.1; -.
SQ   SEQUENCE  221 AA;  24281 MW;  AB324700DE573236 CRC64;

Query Match             15.3%;
Best Local Similarity   41.8%;   Score 353.5;  DB 2;  Length 221;
Matches    76;  Conservative    31;  Mismatches    70;  Indels    5;  Gaps    1;  Pred. No. 4e-15;

Qy  250 ANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDD 309
            :||:|||||||||||| |||| | |  || ||||| ||||| :|:|||||:||||| ::
Db   27 AADKVVHETIIVPKNTTYDGKGQRFVAGKELGDGSQSENQDPVFRVEDGATLKNVVLGAP 86

Qy  310 GADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADT 369
        :||| | || : :|:  ||:|||||:|||    :| | ::| :|||::||||:|:|:|:|
Db   87 AADGVHTYGNVIQNVKWEDVGEDALTVK----KEGKVTIDGGSAQKASDKIFQINKAS 141
```

```
Qy  370 NLSVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFVKSDSEGLNVNTSDISLG 429
          :|  :| :: :|:| || |: ::  ::| : : :::|:: :: ::  :|:|::: ::
Db  142 TFTVKNFTADNGGKFIRQLGGSTFHVDVIIDKCTITNMKEAIFRTDSKTSTVRMTNTRYS 201

Qy  430 DV 431
        :|
Db  202 NV 203
```

```
RESULT 7
Q9X6Z2
ID   Q9X6Z2      PRELIMINARY;      PRT;   222 AA.
AC   Q9X6Z2;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   PECTATE LYASE PRECURSOR (EC 4.2.2.2).
GN   PELA.
OS   Bacillus sp. BP-23.
OC   Bacteria; Firmicutes; Bacillus/Clostridium group;
OC   Bacillus/Staphylococcus group; Bacillus.
OX   NCBI_TaxID=89769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BP-23;
RA   Soriano M., Blanco A., Diaz P., Pastor F.I.J.;
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ237980; CAB40884.1; -.
KW   Signal; Lyase.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26    222       PECTATE LYASE.
SQ   SEQUENCE  222 AA;  23233 MW;  7F29F585791C9682 CRC64;

Query Match             15.1%;
Best Local Similarity   40.9%;   Score 349.5;  DB 2;  Length 222;
Matches    85;  Conservative    35;  Mismatches    71;  Indels    17;  Gaps    6;  Pred. No. 7.1e-15;

Qy  234 SAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAG-SELGDGGQSENQKPL 292
        |||:  ::| | | : :   | |||||:|| ||: |||| | |  ||| |||  ||||| 
Db   10 SAGLVA-SIFGVMPAAAAPTVVNSTIVVPKGTTYDGQGKTFVANPSTLGDGSQAENQKPV 68

Qy  293 FILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITN 352
        | ||:||:|||| :|  ||:||| ||  ::|:| | :| ||||:|:    ::    |:| 
Db   69 FRLEAGATLKNVIIGAPAADGVHCYGSCNISNVVWEDVGEDALTLK--SSGT---VNITG 123

Qy  353 SSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNGGQGNWDLNLSHISAEDGKFSFV 412
         :   : :|: |:|: ::||  ||: :|:|:| || || : | :: |  : ::: : :|
Db  124 GAAYKAYDKVFQMNASGTININKNFRADDIGKLVRQNGGTSYAVNMTLDNSNISNVKDSIM 183

Qy  413 KSDSEGLNVNTSDISLGDVEN--HYKVP 438
        ::|      |:| |: |||    ||:||
Db  184 RTD------SSVSQGKITNTRYSKVP 203
```

```
RESULT 8
Q9EX16
ID   Q9EX16      PRELIMINARY;      PRT;   266 AA.
AC   Q9EX16;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   PUTATIVE SECRETED LYASE.
GN   2SCG38.03.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Saunders D.C., Harris D.;
```

```
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN-A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN-A3(2);
RX   MEDLINE-97000351; PubMed-8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
DR   EMBL; AL445503; CAC13062.1; -.
KW   Lyase.
SQ   SEQUENCE   266 AA;  27499 MW;  7687BFDD93717913 CRC64;

Query Match          9.8%;  Score 226.5; DB 2;  Length 266;
Best Local Similarity 29.6%;  Pred. No. 4.5e-07;
Matches  63;  Conservative  34;  Mismatches  81;  Indels  35;  Gaps   6;

Qy   218 AAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAG 277
             :   ||||||  |  :  |||||   ||
Db    30 SGASSAQAATWPTPNGSEGVSSTLSV------------SGTKDYGMKRLYGTG 70

Qy   278 SELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITV 337
          :| ||||  :|| |  |  ||||   | ||   : ||||||||||||||||||||||||
Db    71 -DLGSGGQDEDQGPILELAPGAVLKNVIIGAPAADGVHCKGSCTLQNVWWEDVGEDAATF 129

Qy   338 KPNSAGKKSHV-EITNSSFEHASDKILQLNADVNLSVDNVKAKDFGTFVRTNGGQQGNW- 395
         |:  ::| |    |:  :|||||||:    |  :|||   |:||||||||||||||
Db,  130 RGSSS---SNVYTVSGGGAKEADDKVFQFNGAGTLNISGFAVNFGTFVRSCGNCSTQYR 186

Qy   396 ---DLNLSHISAEDGKFSFVKSDSEGLNVNTSD 425
            ||  ||:||:    :|:  |  |
Db   187 RTINLNGIEVNWKGGRIA-------GINTNYGD 212


RESULT  9
Q00843
ID   Q00843      PRELIMINARY;      PRT;   215 AA.
AC   Q00843;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE   PECTATE LYASE C.
GN   PELC.
OS   Nectria haematococca mpVI.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Nectriaceae; Nectria.
OX   NCBI_TaxID=70791;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-T8;
RX   MEDLINE-96063610; PubMed-7487098;
RA   Guo W., Gonzalez-Candelas L., Kolattukudy P.E.;
RT   "Cloning of a new pectate lyase gene pelC from Fusarium solani f. sp.
RT   pisi (Nectria haematococca, mating type VI) and characterization of
RT   the gene product expressed in Pichia pastoris.";
RL   Arch. Biochem. Biophys. 323:352-360(1995).
DR   EMBL; U13049; AAA87382.1; -.
KW   Lyase.
SQ   SEQUENCE   215 AA;  22833 MW;  BB0B8EE72B3B0FB5 CRC64;

Query Match          9.4%;  Score 216.5; DB 3;  Length 215;
Best Local Similarity 29.7%;  Pred. No. 1.5e-06;
Matches  62;  Conservative  38;  Mismatches  70;  Indels  39;  Gaps   7;

Qy   220 GGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSE 279
```

```
Db     8 GGVPKPTDH---------ISNS------------KVIEVKAGQVYDGKWAKYDRGSG 43
Qy   280 LGDG-GQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVK 338
          |     |  :| ::|||| :   ||||||||||||||||| || :  || ||||||||
Db    44 ACKGQNEGGDKDAVFLHEGATLKNVIIGKDQSEGVHCKGHCTLEFVWFEDVCEDAISI- 102
Qy   339 PNSAGKKSHVEITNSSFEHASDKIIQLNADTNLSVDNVKAKDFGTFVRT--NGGQGNWD 396
          ||||  ||  :   |  :|:|||||:||:||   ||:|  |:||:|:|     |||| ||
Db   103 ---AGKESW--IIGGGAYHASDKVVQHNGCGTVNIINFYVEDYGKLYRSCGNCSKQCKRN 157
Qy   397 LNLSHISAEDGKFSFVKSDSEGLNVNTSD 425
          :  :  |  : :|::|      :   ||
Db   158 VYIEGVTAKNG------GELAGINANYGD 180


RESULT  10
O53215
ID   O53215      PRELIMINARY;      PRT;   1660 AA.
AC   O53215;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PGRS-FAMILY.
GN   RV2490C OR MTV008.46C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN-H37RV;
RX   MEDLINE-98295987; PubMed-9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; AL021246; CAA16067.1; -.
DR   TubercList; Rv2490c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR000228; -.
DR   InterPro; IPR002173; -.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; -; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
DR   PROSITE; PS01287; RTC; UNKNOWN_1.
SQ   SEQUENCE   1660 AA;  133124 MW;  3A889CE12C0FA945 CRC64;

Query Match          9.0%;  Score 207; DB 2;  Length 1660;
Best Local Similarity 28.1%;  Pred. No. 6.4e-05;
Matches  91;  Conservative  19;  Mismatches 130;  Indels  84;  Gaps  15;

Qy     7 NNNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSP--QSG 64
          |   || :  :|| ||| ||||||||   |   |        |       |::||  :||
Db   476 NGGAGGNGGLV--GDGGAGGHGGDGAAG--------AGYADMTAIFLGSSGTPGEDGG 523
Qy    65 N----AATGAGG-NDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQA--IT 117
          |     ||| ||  :  |||  | |||  |:          |   |:||||| || || ||
Db   524 NGGAGGAGGAHAGDGGAGGNGGAGGNGAHGFNAVLVSDGGNGGDGGAGGRG 583
Qy   118 PDGQGGQGIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNNSASSGTSSSGG---SPFND 174
          | |  || |  |                 |   |    |   | : |:||     || |
Db   584 GDGGAGGAGGDAPAGRA---------GSQVGGDGGAGGAGGGGSGGRGRDMAFKD 633
Qy   175 LSGGKAPSGNSP-SGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVG 233
```

```
Db   634 GDGGAGGDGGDPGAGGKGGAG-------------------GAGATEGVTGATGATVHSG--G 674
             ||                                   |:|  |:|
Qy   234 SAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLF 293
              :|:||  ||||                                         |||
Db   675 NGGKG-GNGADATVAGAN------------------GGKGGAGNGGLVGDGGAG------- 710
Qy   294 ILEDG-----ASLKNVTMGDDGADG 313
              |||       ::|||||
Db   711 --GDGGSGAAGANGANVGEDGADG 732
```

RESULT 11
O53775
ID   O53775;   PRELIMINARY; PRT; 1306 AA.
AC   O53775;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   PGRS-FAMILY PROTEIN.
GN   RV0578C OR MTV039.16C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; AL021942; CAA17449.1; -.
DR   TubercuList; Rv0578c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR000209; -.
DR   InterPro; IPR002173; -.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; -; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_2.
DR   PROSITE; PS00138; SUBTILASE_SER; UNKNOWN_1.
SQ   SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;

Query Match   8.6%; Score 199.5; DB 2; Length 1306;
Best Local Similarity 26.0%; Pred. No. 0.00014;
Matches 88; Conservative 16; Mismatches 122; Indels 113; Gaps 11;

Qy   8 NNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAA 67
             : : | ||| ||| || |||
Db   754 NGVAGSGQPGAGGDGGDGGTGGVGVGNGGRGIDGADGAT------------------- 789
Qy   68 TGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGGQIG 127
             | |: |||| |||||||| |
Db   790 --AGARGQDGGAGGAGGAGKGGRGGTGGP------------GGAGPAGTTGSQGAGGNG 832
Qy   128 DNPLLKAMLKLIARMMDGQSDQFGQPGTPGNNSASSGTSSSSGGSPFNDLSGGKA-PSGNSP 186
             ||| |:||| || || |||||| || | | |||
Db   833 ------------GSGGTGGDPGDGGNGANGSVFTNNGIGGNGGNGGNAGPSGAGG 875
Qy   187 SGNYSPVSTFSPPSTPTSPTSPL-----------DFPSSPTKAAG----------G 221
             ||| ||| :||| |
Db   876 SGGAG--STFG----ATGSSSSIHVNGNGNGGNGDHALSGNGAAGGNGGNGNGSLRG 929
Qy   222 STPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELG 281
```

```
Db   930 SGGAGGHGGNGGNGGNASRGMGGDGGTGGAGGN------------AQGIGNG----GAGGNGG 973
             |||  |                         |           |||:|:|       |
Qy   282 DGGQSENQKPLFIL------EDGASLKNVTMGDDGADG 313
             |||  |                       |  ::  ||||
Db   974 DGGTGSDGNPGAITGSGGRGGDGGVGGGGSVAGDGADG 1012
```

RESULT 12
Q00845
ID   Q00845;   PRELIMINARY; PRT; 233 AA.
AC   Q00845;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PECTATE LYASE D.
GN   PELD.
OS   Nectria haematococca mpVI.
OC   Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC   Hypocreales; Nectriaceae; Nectria.
OX   NCBI_TaxID=70791;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=T8;
RX   MEDLINE=96400363; PubMed=8806739;
RA   Guo W., Gonzalez-Candelas L., Kolattukudy P.E.;
RT   "Identification of a novel pelD gene expressed uniquely in planta by
RT   Fusarium solani f. sp. pisi (Nectria haematococca, mating type VI) and
RT   characterization of its protein product as an endo-pectate lyase.";
RL   Arch. Biochem. Biophys. 332:305-312(1996).
DR   EMBL; U13050; AAC49420.1; -.
KW   Lyase.
SQ   SEQUENCE 233 AA; 24509 MW; 2FF53A0D9A4E2CB9 CRC64;

Query Match   8.4%; Score 194.5; DB 3; Length 233;
Best Local Similarity 29.9%; Pred. No. 3.9e-05;
Matches 50; Conservative 34; Mismatches 74; Indels 9; Gaps 3;

Qy   260 TVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGD 319
             |:||::||| | ::||| : | |::|| :|| :| ||:| |||
Db   40 TLKKGEVFDAGWVRYDRGVKCSGQAEGGSKDAVFILEEGATLRNVIIGANQREGIHCKGS 99
Qy   320 AKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLSVDNVKAK 379
             : ||||| |||||| ::: ||||||| |||| : : | | :|
Db   100 CNIEFAWFEDVCEDAISILGSGT-----ANIIGGGAYHASDKVIQHNGCGHVNIVNFYAN 154
Qy   380 DFGTFVRTNGGQQGNWDLNLS-HISAEDGKFSFVKSDSEGLNVNTSD 425
             | | || || | | : |:| ||| |:| | ||| || |
Db   155 DYGKVYRSCGNCKGNTNCKRSVHM--EGTTAVKGGELIGINTNYGD 198
```

RESULT 13
O93877
ID   O93877;   PRELIMINARY; PRT; 240 AA.
AC   O93877;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE   PECTATE LYASE.
OS   Fusarium oxysporum f. sp. lycopersici.
OC   Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Fusarium.
OX   NCBI_TaxID=59765;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=42-87;
RA   Huertas-Gonzalez M., Ruiz-Roldan M., Garcia-Maceira F., Roncero M.,
RA   Di Pietro A.;
RT   "Cloning and characterization of pll encoding an in planta expressed
RT   pectate lyase of Fusarium oxysporum.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF080485; AAC64368.1; -.
KW   Lyase.

```
SQ   SEQUENCE   240 AA;   24859 MW;   46D4B29730S5006B1 CRC64;

Query Match        8.4%;   Score 193;   DB 3;   Length 240;
Best Local Similarity    31.5%;   Pred. No. 5e-05;
Matches    64;   Conservative    31;   Mismatches    76;   Indels    32;   Gaps    9;

Qy   235  AGIGAGNSVAFT-----SAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQ  289
          :| ||.|||:||      | |  :   || :: :|||| .| .| ||| :| :||
Db    10  AGV--SSAAVTKTLPKSAGA--TSFPTAVPVKGS--YDGGWKRFERSPSVCQGSETGE    62

Qy   290  K-PLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHV  348
          : ||: | | : :: || : ||  | | ||.| |   |||.| |  :
Db    63  KDAMFILENGATLSNVIIGASOAEGVHCKGTCTLNNVWADVCEDAVTLKQTSGTS----  118

Qy   349  EITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRT------NGGQQGNWDLNLSHI  402
          :| | :|| ||. :  |:: | :: |||  :| : :  |      ||  |   :| |
Db   119  YINGGGAFHASDKIVQFNGRGTVQIKDFYAEDYGKLVRSCGNCKDNGGPR---NVVISGS  175

Qy   403  SAEDGKFSFVKSDSEGLNVNTSD  425
          || |  :|| .| |||
Db   176  VAVDGGVLC------GINTNYGD  192


RESULT   14
ID   O53212             PRELIMINARY;       PRT;    694 AA.
AC   O53212;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE   PGRS-FAMILY PROTEIN.
GN   RV2487C OR MTV008.43C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; AL021246; CAA16064.1; -.
DR   TubercuList; Rv2487c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR002173; -.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; -; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; UNKNOWN_1.
SQ   SEQUENCE   694 AA;   57218 MW;   7D4387928306BC51 CRC64;

Query Match        8.4%;   Score 193;   DB 2;   Length 694;
Best Local Similarity    25.4%;   Pred. No. 0.00017;
Matches    86;   Conservative    28;   Mismatches   140;   Indels    84;   Gaps   14;

Qy    18  QSGGDNGLGGH-NANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGGNDQT   76
          |:| || || | | |:| :  | ||     |||| : :| | | | | | | : :  |
Db   256  QTCGDGGTGGHGTAGTPTGGTGGDGAT-----------------ATAGSKATGGAGGDGGT  300

Qy    77  TGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMCGNNGLDQAITPDGQGGQIGDNPLLKAML  136
          || | | || | ||:| : :
Db   301  AAAGGGGGGNGGDGGVA----QGDIASAFGGDGGNGSDVVAAGSGGGGSGAGGG----AFV  352
```

```
Qy   137  KLIARMDGQSDFGQPGTGNNSASSGT-SSSGGSPFNDLSGGK--APSGNSPSGNYSPV  193
          :| | : | :  | | |: | :   | .| | ||  |  | |    | |  | |
Db   353  HIATATSTGGSGFG--GNGAASAASGADGGAGGDGGAGGGLLFGDGGNGGAGGAGGI   410

Qy   194  STFSPPSTPTSPTSPLDFSPSSPTKAAGGSTPVT-DHPDP----------VGSAG  236
          : :  |  | |    | |  |  |  | |  | :  :||
Db   411  GG------------DGATGGPGGSGGNAGIARFDSPDPEAEPDVVGGKGGDGKGGSG  456

Qy   237  IGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFILE  296
          :| | |  | :| | | :   |   | ||. |  |    | |   | || |
Db   457  LGVGGAGGTGGAGGNG--------GAGGLLFGNGGN--GGNAGAGG------------D  493

Qy   297  DGASLKNVTMGDDGADGIHLYGDAKIDNL-HVTNVGED  333
          .| |   | :|  || | :: | |   : :||
Db   494  GGAGVAGGVGVGNGGGGGGTATFHEDPVAGWAVGGVGGD  531


RESULT   15
ID   Q9NBW0             PRELIMINARY;       PRT;    929 AA.
AC   Q9NBW0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   BRAKELESS-A.
GN   SBB OR BKS OR CG5580.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20265906; PubMed=10804172;
RA   Senti K., Keleman K., Eisenhaber F., Dickson B.J.;
RT   "brakeless is required for lamina targeting of R1-R6 axons in the
RT   drosophila visual system.";
RL   Development 127:2291-2301(2000).
DR   EMBL; AF242193; AAF76321.1; -.
DR   FlyBase; FBgn0010575; sbb.
DR   InterPro; IPR000255; -.
DR   PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN_1.
SQ   SEQUENCE   929 AA;   88820 MW;   5F1BA2B89187ABF CRC64;

Query Match        8.3%;   Score 191;   DB 5;   Length 929;
Best Local Similarity    21.7%;   Pred. No. 0.00033;
Matches    96;   Conservative    51;   Mismatches   175;   Indels   120;   Gaps   12;

Qy     7  NNNTSSSPGLFQS-GGDNGLGGHNANSALGQQPIDRQTIEQMAQL-  50
          ||| :: |  :|: .| || || | |:| ||  | ||
Db   416  NNNSTSAGGPPSSQQGNNGNGSGSSSSSGKSSAKMSIDHQATLDKGLKMKIKRTKPGT  475

Qy    51  ----LAELLKSLLSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLS  105
              :|  : |:|    | |  | | | |  ||  | |
Db   476  KSSEAKHEIVKA--TDQQQNGALGAGSNNSANEDGSGSGSSTNASSLGSTNSSSSASSGS  533

Qy   106  EMGNNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMDGQSDQFGQPGTGNNSASSGTS  165
          : | | :  |     |:
Db   534  SSSS---------------------------GSSSSSSKKHLNNASSGSGSS  558

Qy   166  SSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPS---TPTSPTSPLDFSSPTKAAGG  221
          |||| ::| .| |  | | : |  :  :| :|  |   | |||  | |||  || ||
Db   559  SSGGGSQNNASGHASGGGSSGSSGSSSQSTPQGTKRGSSGHRREKTKDNAHSNRMSVDKSAAA  618

Qy   222  STPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELG  281
          ::  ::     ||:| || |  | :| || |  |  ||:  | | | |  : | ||
Db   619  ASAAGEKDTPEKCSGTGAGGSPCSCNGDVGAPCSHHACIRRAAHMSNSAG--NANSSAG  675

Qy   282  DG-----------------------GQSENQKPLFILEDGASLKNVT------MGDDGAD  312
          .|                        : | :|  : : :|| |:: :      |||| |
Db   676  TGQSGGSSMSAVPPGVFTPSAGSPSTGSPSTVVPAAASLLAATGAASSSASQMASSSAG  735
```

```
Qy   313  GIHLYGDAKIDNLHVTNVGEDAITVKPNSAGKKSHVEITNSSFEHASDKILQLNADTNLS  372
          |:  |         | :|     |     |:  | | |  | : :  :    :  |
Db   736  GVGGSGG-------------GANA----PGPGKESAGSIKISS--HIAAQLAAAASNSYS  778

Qy   373  VDNVKAKDFGTFVRTNGGQQGN  394
          |:  | ||| |
Db   779  --------GSGANTNQGQNSN  791
```

Search completed: October 18, 2001, 09:10:32
Job time: 106 sec

```
                    Copyright (c) 1993 - 2000 Compugen Ltd.
                    GenCore version 4.5

OM protein - protein search, using sw model

Run on:    October 18, 2001, 09:08:46 ; Search time 13.15 Seconds
                                         (without alignments)
                                         1164.426 Million cell updates/sec

Title:         US-09-596-958-2
Perfect score: 2310
Sequence:      1 MSILTLNNTSSSPGLFQSG.........LGDVENHYKVPMSANLKVAE 447

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters:     93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :  SwissProt_39:*

    Pred. No. is the number of results predicted by chance to have a
    score greater than or equal to the score of the result being printed,
    and is derived by analysis of the total score distribution.
```

SUMMARIES

Result NO.	Score	% Query Match	Match Length	DB	ID	Description	
1	177.5	7.7	1901	1	YZ08_MYCTU	O53553	mycobacteri
2	171	7.4	398	1	CSP_PLACC	P08673	plasmodium
3	171	7.4	463	1	YA68_MYCTU	O53416	mycobacteri
4	169.5	7.3	914	1	WA22_MYCTU	O06794	mycobacteri
5	166	7.2	603	1	YD25_MYCTU	O10637	mycobacteri
6	165	7.1	959	1	N100_YEAST	Q02629	saccharomyc
7	163	7.1	400	1	RTOA_DICDI	P54681	dictyosteli
8	162	7.0	434	1	YK98_MYCTU	O10707	mycobacteri
9	160.5	6.9	314	1	YK83_MYCTU	O10691	mycobacteri
10	160.5	6.9	937	1	HYR1_CANAL	P46591	candida alb
11	160.5	6.9	957	1	Y278_MYCTU	P56077	mycobacteri
12	159.5	6.9	778	1	YQ34_MYCTU	P71933	mycobacteri
13	157.5	6.8	2329	1	YS89_CAEEL	Q09624	caenorhabdi
14	157	6.8	543	1	YP91_MYCTU	Q50630	mycobacteri
15	155	6.7	1113	1	N116_YEAST	Q02630	saccharomyc
16	155	6.7	2249	1	190K_RICRI	P15921	rickettsia
17	154	6.7	403	1	HRPN_ERWAM	Q01099	erwinia amy
18	154	6.7	430	1	NU42_YEAST	P49686	saccharomyc
19	153	6.6	788	1	SP1_RAT	Q01714	rattus norv
20	153	6.6	1049	1	CA13_BOVIN	P04258	bos taurus
21	152.5	6.6	2003	1	YDBA_ECOLI	P33666	escherichia
22	150.5	6.5	598	1	YAUF_SCHPO	Q10168	schizosacch
23	150	6.5	515	1	YI40_MYCTU	Q50594	mycobacteri
24	148	6.4	801	1	Y747_MYCTU	O53810	mycobacteri
25	147	6.4	1453	1	CA11_CHICK	P02457	gallus gall
26	146.5	6.3	366	1	CAS4_EPHMU	P18503	ephydatia m
27	146	6.3	937	1	NU98_RAT	P49793	rattus norv
28	145	6.3	1736	1	CA2B_HUMAN	P13942	homo sapien
29	144	6.2	1156	1	GLH4_CAEEL	O76743	caenorhabdi
30	144	6.2	1581	1	PPRB_HUMAN	Q15648	h peroxisom
31	141.5	6.1	419	1	CSP_PLACM	P08676	plasmodium
32	141.5	6.1	541	1	NU57_YEAST	P48837	saccharomyc
33	141	6.1	1150	1	APMU_PIG	P12021	sus scrofa
34	140.5	6.1	2132	1	PGCA_MOUSE	Q61282	mus musculu
35	140	6.1	671	1	HMOC_DROME	P22810	drosophila
36	140	6.1	678	1	YF48_MYCTU	Q10778	mycobacteri
37	139	6.0	1325	1	YDEK_ECOLI	P32051	escherichia
38	138.5	6.0	630	1	MUC1_MOUSE	Q02496	mus musculu
39	138.5	6.0	1475	1	N153_HUMAN	P49790	homo sapien
40	138.5	6.0	1516	1	NCO2_XENLA	Q9w705	xenopus lae
41	138	6.0	671	1	CA11_RAT	P02454	rattus norv
42	138	6.0	2090	1	N214_HUMAN	P35658	homo sapien
43	138	6.0	3591	1	FHAB_BORPE	P12255	bordetella
44	137.5	6.0	283	1	YO33_CAEEL	Q09233	caenorhabdi
45	137.5	6.0	573	1	C114_MOUSE	P19467	mus musculu

ALIGNMENTS

```
RESULT 1
YZ08_MYCTU
ID   YZ08_MYCTU     STANDARD;      PRT;   1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV3508 PRECURSOR.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98995987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL022022; CAA17745.1; -.
DR   HSSP; P19972; 1KVD.
DR   TuberculoList; Rv3508; -.
DR   InterPro; IPR000084; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1901       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV3508.
SQ   SEQUENCE   1901 AA;  147627 MW;  C7B1923D5D0146CD CRC64;

Query Match       7.7%;  Score 177.5;  DB 1;  Length 1901;
Best Local Similarity   25.9%;
Matches   79;  Conservative   22;  Mismatches 109;  Indels   95;  Gaps   11;
                                                       Pred. No. 0.049;

Qy    20 GGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGGNDQTTGV 79
```

```
Db   599  GGSSGAGGTNGSGGAGG------              ----------TGGQQGGAGGAGGAGGADNPTGI 636
                 ||:||:||: |                               ||  |||| ||| |||:
Qy    80  GNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGG--GQIGDNPLLKAMLK 137
                 | ||| ||  |                     -GGAGGAIGTGGTGGAVGSVGNAGIGGT--- 678
Db   637  GGAGGTGGTGGAGGA--------
                 | || |||  |||                    :||:  :
Qy   138  LIARMMDGQSDQFGQPCTGNNSASSGTSSSGGSFNDLSGGK-APSGNSPSGNYSPVSTF 196
                 :|:::::|::|:  ||                     ::||
Db   679  -----GGTGGVGGAGGAGAAAAGSSATGGAGFAGGAGGCGAGGNSGVG--------- 723
                 :     || ||| |                      :|||  ||
Qy   197  SPPSTPSPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAG------IGAGNSVAFTSA 248
                 :  :   |||| ||                       ::| |
Db   724  --GTNGSGGAGGAGGKGGTGGAGGSG--ADNPTGAGFAGGAGGTGGAAGAGGAGGATGT 778
                 | ::|| ||||| ||                      ||| |||| |||
Qy   249  GANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGQSENQKPLFILEDGASLKNVTMGD 308
                 | |: :||                            ||||   ||
Db   779  GGTGGVVGAT-------GSAGIGAGGAGGGGGDGG------DGAS--GLGLGL 814
                 |:                                  |||||||
Qy   309  DGADG 313
                 ::||
Db   815  SGFDG 819
```

RESULT 2
CSP_PLACC
```
ID   CSP_PLACC      STANDARD;       PRT;   398 AA.
AC   P08673;
DT   01-JAN-1988 (Rel. 06, Created)
DT   01-JAN-1988 (Rel. 06, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   CIRCUMSPOROZOITE PROTEIN PRECURSOR (CS).
OS   Plasmodium cynomolgi (strain Ceylon).
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5829;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=87102878; PubMed=3802196;
RA   Galinski M.R., Arnot D.E., Cochrane A.H., Barnwell J.W.,
RA   Nussenzweig R.S., Enea V.;
RT   "The circumsporozoite gene of the Plasmodium cynomolgi complex.";
RL   Cell 48:311-319(1987).
CC   -!- MISCELLANEOUS: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT
CC       SURFACE ANTIGEN ON THE SPOROZOITE (THE INFECTIVE STAGE OF THE
CC       MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE
CC       VERTEBRATE HOST).
CC   -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR
CC       ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES
CC       WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M15103; AAA29533.1; -.
DR   PIR; C26255; O2ZOAS.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR003067; -.
DR   Pfam; PF00090; tsp_1; 1.
DR   PRINTS; PR01303; CRCMSPRZOITE.
KW   Malaria; Sporozoite; Repeat; Signal.
FT   SIGNAL       1     19       PROBABLE.
FT   CHAIN       20    398       CIRCUMSPOROZOITE PROTEIN.
FT   DOMAIN      97    249       17 X 9 AA TANDEM REPEATS OF A-G-N-N-A-A-
FT                               A-G-E.
SQ   SEQUENCE   398 AA;  37718 MW;  6DFA2E8A62ED05BF CRC64;
```

```
Query Match            7.4%;  Score 171;  DB 1;  Length 398;
Best Local Similarity 24.4%;  Pred. No. 0.019;
Matches  89;  Conservative  .46;  Mismatches 162;  Indels  68;  Gaps  15;

Qy     4  LTLNNTSSSPGLFQ----SGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLL 59
               :: :||   ||                    :  |  | :  |
Db    37  VSFNNVDASSLGAAQVRQSASRGRGLGENPKNEEGADKP--KKKDEKQVEPKKPRENKLK 94
               :|| || |                           |   :|
Qy    60  SPQSGNAATGAGGNDQTTG-VGN---AG-GLNGRKGTAGTTP--QSDSQNMLSEMGNN 110
               |||:|| |                            :|: | :
Db    95  QPAGNNAAAGEAGNNAAAGEAGNNAAAGEAGNNAAAGEAGNNAAAAGEAGNN 154
               |:: | ||                            | :::::||
Qy   111  GLDQAITPDGQGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPG------TGNNSAS-- 161
               |   |  | :  |                       :|  | :::|:::
Db   155  AACGA-AGNNAACGEAGNN--------AAGGAAGNNAAAGEAGNNAAAGEAGNNAAAAGE 204
               | :  :||| ||                        :|  ||| ||:
Qy   162  SGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPSPTSPLDFPSSPTKAAGG 221
               :|  |||| ||  |                      |   :|  ||
Db   205  AGNNAAGGAAGNNAAAGEA--GNNAAAG------- 239
               :|||| ||  ||                        :||||||
Qy   222  STPVTDHPDPVGSAGIG-AGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSEL 280
               |:|  | |: :  |||                     :  :|::||| |
Db   240  AAGNNAAAGEAGAGGAGRAGNNAAAAGEAGAGGAGRAGN-NAAAGEAGAGGAGNAGNKKA 298
               |   |  || |                         :||| |
Qy   281  GDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPN 340
               | || | :  |                         : ||    ||
Db   299  GDAGQGQN------NGGANVPNVKLVKEYLDKIRSTIGVEWSPCSVTCGKGVRMRRKVN 351
               |  |:|                                |
Qy   341  SAGKK 345
               :|:||
Db   352  AANKK 356
```

RESULT 3
YA68_MYCTU
```
ID   YA68_MYCTU     STANDARD;       PRT;   463 AA.
AC   O53416;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV1068C.
GN   RV1068C OR MTV017.21C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Sqares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021897; CAA17184.1; -.
```

DR   TubercuList; Rv1068c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR002952; -.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR01228; EGGSHELL.
KW   Hypothetical protein.
SQ   SEQUENCE   463 AA; 39305 MW; CF5696A7E9593952 CRC64;

Query Match   7.4%; Score 171; DB 1; Length 463;
Best Local Similarity 25.8%; Pred. No. 0.022;
Matches   90; Conservative   28; Mismatches 121; Indels 110; Gaps 17;

Qy   6 LNNNTSSSPGLFQSGSGGDNGLGGHNAN---------------SALGQQPIDRQTIEQMA 48
Db   150 LTGGTGGSAGLICNGNGGAGGANGGAGGNGGWLYCSGGNGGAGGAG---------- 201

Qy   49 QLLAELLKSLLSPQSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMG 108
             ::|||| |:|||| |:||| |||| || |
Db   202 ------AIGAPGVAGGAGGTAGLFGNGGAGGAGGAGGAGGRGDGGSAGWLS--G 251

Qy   109 NNGLDQAITPDGQGGQGIGDNPLLKAMLKLIARMDGQSDQF--GQPGTGNNSASSGTSS 166
             ::|-| ||| ||||| ||| :| :|:
Db   252 NGG--DAGT--GGGGNAGNGG----------------NGGSAGWLSGNGGTGGGGTAGAGG 294

Qy   167 SGG--SPFNDLSGGK-APSGNSPSGNY--SPVSTFSPPSTPTSPTSPLDFPSSPTKAAG 220
             ||| | | :|: |
Db   295 QGGNGNSGIDPGNGGQGADTGNAGNGGHGGSAAKLFG------------------DGGAG 336

Qy   221 GSTPTVDHPDPVGSAGIG------AGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTF 274
             |:|: | | |
Db   337 GAGGMGSTGGTGGGGFGGTGGGTGNGGNHAGGAGGSGGT------AGLL--GSGGSG 385

Qy   275 TAGSELGDGGQSENQKPLFILEDGASLK-NVTMGDDGADGIHLYGDAKI 322
             ||:|||||||| ||||||||| |||||: ||||:
Db   386 GTGDGGNGG--------------------LGAGSGAKGNGNGGDGGKG----GDAQL 420

RESULT   4
WA22_MYCTU
ID   WA22_MYCTU   STANDARD;   PRT;   914 AA.
AC   O06794;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation.update)
DE   WAG22 ANTIGEN PRECURSOR.
GN   WAG22 OR RV1759C OR MTCY28.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC   SUBFAMILY.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way

CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z95890; CAB09322.1; -.
DR   HSSP; P41140; 2SFA.
DR   TubercuList; Rv1759c; -.
DR   InterPro; IPR000084; -.
DR   Pfam; PF00934; PE; 1.
KW   Antigen; Repeat; Signal.
FT   SIGNAL   1   30   POTENTIAL.
FT   CHAIN   31   914   WAG22 ANTIGEN.
SQ   SEQUENCE   914 AA; 74354 MW; F6953C3DB8E6ACB CRC64;

Query Match   7.3%; Score 169.5; DB 1; Length 914;
Best Local Similarity 26.7%; Pred. No. 0.057;
Matches   91; Conservative   23; Mismatches 136; Indels 91; Gaps 16;

Qy   6 LNNNTSSSPG------------LFQSGSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELL 55
            ||||||||| | :|||| |:|||| |
Db   532 IGNGDSGTPGTGDDGAGGWLFGNGNGGAGGAAGTNGSAG------------ 571

Qy   56 KSLLSPQSGNAATGAGG-NDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQ 114
            | |||| :|:|||| |||||||| ||||||||||||||| |:||||||
Db   572 ---------GAGGAGGILFGTGAGGAGGV-GTAGAGGAGGAGGSAFLIGSGGTGGVGG 620

Qy   115 AITPDQGGQGGGIGDNPLLKAMLKL-------IARMDGQSDQFGQPGTGNNS---ASSG 163
            |: | | || |: ||| :|:| |
Db   621 AATTTGGVGGAGGNAGLLIGAAGLGGCGGGAFTAGVTTG----GAGGTGGAAGLFANGG 675

Qy   164 TSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSP----PSTPTSPTSPLDFPSSPTKAA 219
            | ||||| | |:||| | | ||| |:
Db   676 AGGAGGT--GSTAGGAGGAGGA-GGLYAHGGTGGPGGNGGSTGAGGTGAGGPGG-LYGA 731

Qy   220 GGSTPTVDHPDPVGSAGIGAGNSVAFT--SAGANQTVLHDTITVKAGQVFDGK-----GQ 272
            |||| |: || :| | ||| :|:|||| |||||||| || |
Db   732 GGSGGAGGHGGMAGGGGGVGGNAGSLTLNASGGAGGS---------GGSSLSGKAGAGGA 782

Qy   273 TFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADG 313
            | ||| | |:| | ||||||| ||:|||||
Db   783 GGSAGLFYGSGGAGGN---------------GGYSL-NGTGDGGTGG 814

RESULT   5
YD25_MYCTU
ID   YD25_MYCTU   STANDARD;   PRT;   603 AA.
AC   Q10637;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV1325C PRECURSOR.
GN   RV1325C OR MTCY130.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS

```
CC   SUBFAMILY.
CC   ------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------
DR   EMBL; 273902; CAA98089.1; -.
DR   HSSP; P19972; 1KVD.
DR   TubercuList; Rv1325c; -.
DR   InterPro; IPR000084; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    603       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV1325C.
FT   DOMAIN      114    603       GLY-RICH.
SQ   SEQUENCE   603 AA;  49575 MW;  4F9BCB82B07AE964 CRC64;

Query Match             7.2%;  Score 166;  DB 1;  Length 603;
Best Local Similarity  24.9%;  Pred. No.: 0.054;
Matches   83;  Conservative  15;  Mismatches 130;  Indels 106;  Gaps  12;

Qy    10  TSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATG   69
              |||  ||  |||  |  : |  |                |:    ||
Db   274  TGGGGGLFSNGGAGGAGGFGVSGSAG----------------------------GNGGTG  305

Qy    70  A------GGNDQT--TGVGN--AGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAIT  117
                    ||||||  |||    || |  |  ||   :||  |      :  |  ||
Db   306  GDGGIFTGNGGTGGTGGTGTGNQLVGGEGGAGGAGGNAGILFGAGGIGGTGGTGLG-APD  364

Qy   118  PDGQGG-GQIQDNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLS  176
             | | |    |  |          :| ||  :   |  |||    ||     :   :
Db   365  PGGTGCKGGVGG--------------IGGAGALFGPGGAG------GTGGFGASSADQMA  404

Qy   177  GGKAPSGNS----------PSGNYSPVSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPV  225
             ||   || |         |  :  ||         |   ||||  ||
Db   405  GGIGGSGGSGGAAKLIGDGGAGGTGDSVRGAAGSGGTGGTGGL----IGDGGAGGAGGT  460

Qy   226  TDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQTTFAGSELGDGGQ  285
             ||  | |||  |||    |  ||       |     |  ||  |  |     |
Db   461  GIEFGSVGGAGGAGGNAAGLSGAG----------------------GAGGAGGFGETAGDGGA  501

Qy   286  SENQKPLFILEDGASLKNVTMGDDGADGIHLYGD  319
             |||  |  |||||||||||||||  | || ||
Db   502  GGN----------AGLLNGDGGAGGAGGLGIAGD  525

RESULT   6
ID   N100_YEAST     STANDARD;      PRT;    959 AA.
AC   Q02629;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   NUCLEOPORIN NUP100/NSP100 (NUCLEAR PORE PROTEIN NUP100/NSP100).
GN   NUP100 OR NSP100 OR YKL068W OR YKL336.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93054906; PubMed=1385442;
RA   Wente S.R., Rout M.P., Blobel G.;
RT   "A new family of yeast nuclear pore complex proteins.";
RL   J. Cell Biol. 119:705-723(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=94378724; PubMed=8091863;
RA   Rasmussen S.W.;
RT   "Sequence of a 20.7 kb region of yeast chromosome XI includes the
RT   NUP100 gene, an open reading frame (ORF) possibly representing a
RT   nucleoside diphosphate kinase gene, tRNAs for His, Val and Trp in
RT   addition to seven ORFs with weak or no significant similarity to
RT   known proteins.";
RL   Yeast 10:S69-S74(1994).
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC       NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLOCATION OF
CC       THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC   -!- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC   ------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------
DR   EMBL; 215035; CAA78753.1; -.
DR   EMBL; X75780; CAA53406.1; -.
DR   EMBL; 228068; CAA81905.1; -.
DR   PIR; B44402; B44402.
DR   PIR; S39173; S39173.
DR   PIR; S44518; S44518.
DR   SGD; S0001551; NUP100.
KW   Nuclear protein; Transport; Repeat.
FT   DOMAIN       33    571       29 X 6 AA APPROXIMATE REPEATS OF
FT                                G-L-F-G.
SQ   SEQUENCE   959 AA;  99988 MW;  D3985F99O1BBAA51 CRC64;

Query Match             7.1%;  Score 165;  DB 1;  Length 959;
Best Local Similarity  22.9%;  Pred. No. 0.1;
Matches  117;  Conservative  51;  Mismatches 198;  Indels 146;  Gaps  23;

Qy     5  TLNNNTSSSPGLF--------QSGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKS   57
          ||:  |||                 :|| |||| ||||| ||||||||
Db   261  TLGNTTNNRNGLFGQMNSSNQGSSNSLFGQNSMNS-STQGVFGQNNQM-QINGNNNNS  318

Qy    58  LLSPQS--GNAATGA-----------GGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQ  101
          ||:|:                    |  ||| ||| || ||| |  ||
Db   319  LFGKANTFSNSASGGLFGQNNQQGGSGLFGQNSGTSGSSGLFGQNNQK-QPNTFTQSNTG  377

Qy   102  NMLSEMGNNGLDQA---------ITPDGQGGQIQDNPLLKAMLKLIARMMDGQSDQFGQP-  153
           | ||| ||                  | |||  | |||   |  |     |  || |
Db   378  IGLFGQNNQQQQSTGLFGAKPAGTTGSLFGGN------------------SSTQPNSL  418

Qy   154  -GTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPSGNYSPVSTFSPPSTPTSPTSPLDFP  212
             ||| |  |||  || |   |  ||     ||
Db   419  FGTTNVFTSNTQSQQGNSLFGATKLTNMPFGGNPTANQSGSGNSLFGTKPASTTGSL---  475

Qy   213  SSPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKACQVFDGKGQ  272
              | ||  |   |
Db   476  ------FGNNTASTTVPSTNGLFGNNANNSTSTNTGL-----------FGAKPDSQSK  517

Qy   273  TFTAGSELGDGGQSEN---QKPLFILEDGASLKNVTMGDDGADGIH----------LYGD  319
                |  | |               | || || ||||   | |||| ||
Db   518  PALGGGLFGNSNSNSSTIGQNKPVF----GGTTQNT-GLFGATGTNSSAVGSTGKLFG-  570

Qy   320  AKIDNLHVTNVGEDAITVKPNSAGKK---------SHVEITNSSFEHASDKILQLNADT  369
               | | |  |            ||
Db   571  ---QNNNTLNVGTQNVPPVNNTTQNALLGTAVPSLQQAPVTN--EQLFSKISIPNSIT  624

Qy   370  NLSVDNVKAKDFGTFVRTNGGQQGNWDLNLSH---------ISAEDGKFSFVK-----  413
                ||     |            |   | |       ::
Db   625  N----PVKA----TTSKVNADMKRNSSLTSAYRLAPKLFAPSSNGDAKFQKWGKTLERS  676

RN   [1]
RP   SEQUENCE FROM N.A.

RN   [2]
RP   SEQUENCE FROM N.A.
```

```
Qy    414 ------------SDSEGLNVNTDISLGDVENHY 435
                      :| | .:|.:|:.| | .|
Db    677 DRGSSTSNSITDPESSYLNSNDL-LFDPDRRY 707


RESULT 7
RTOA_DICDI
ID   RTOA_DICDI    STANDARD;    PRT;    400 AA.
AC   P54681;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   RTOA PROTEIN (RATIO-A).
GN   RTOA.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97108762; PubMed=8951083;
RA   Wood S.A., Ammann R.R., Brock D.A., Li L., Spann T., Gomer R.H.;
RT   "RtoA links initial cell type choice to the cell cycle in
RT   Dictyostelium.";
RL   Development 122:3677-3685(1996).
CC   -!- FUNCTION: MAY HAVE DUAL FUNCTIONS, ONE IN VEGETATIVE CELLS AT
CC       STARVATION INVOLVING CELL-TYPE CHOICE AND A LATER FUNCTION
CC       INVOLVING PROGRESSION TO THE TIPPED MOUND STAGE.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED IN VEGETATIVE CELLS. LEVELS
CC       DECREASE DRAMATICALLY AS DEVELOPMENT BEGINS, AND REMAIN LOW
CC       THROUGHOUT THE AGGREGATION AND LOOSE MOUND STAGES. THE LEVELS
CC       BECOME MORE ABUNDANT AGAIN AT THE TIGHT-MOUND STAGE AND REMAIN
CC       HIGH THROUGHOUT THE REST OF DEVELOPMENT AND FRUITING BODY
CC       FORMATION.
CC   --------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; U48298; AAC47311.1; -.
DR   DictyDb; DDO2056; rtoA.
KW   Transmembrane; Repeat.
FT   TRANSMEM     30     50       POTENTIAL.
FT   DOMAIN       30     38       POLY-LEU.
FT   DOMAIN       87     96       POLY-SER.
FT   DOMAIN      181    181       11 X 11 AA TANDEM REPEATS OF G-S(2)-N-S-
FT                               G-S-E-S(3).
FT   REPEAT      181    190       1 (INCOMPLETE).
FT   REPEAT      191    200       2 (INCOMPLETE).
FT   REPEAT      201    211       3.
FT   REPEAT      212    222       4.
FT   REPEAT      223    233       5.
FT   REPEAT      234    244       6.
FT   REPEAT      245    255       7.
FT   REPEAT      256    266       8.
FT   REPEAT      267    277       9.
FT   REPEAT      278    287      10 (INCOMPLETE).
FT   REPEAT      288    298      11.
SQ   SEQUENCE   400 AA;  39840 MW;  0392F4E68DC27A75 CRC64;

Query Match             7.1%;  Score 163;  DB 1;  Length 400;
Best Local Similarity   21.8%;  Pred. No. 0.049;
Matches   83;  Conservative   54;  Mismatches  151;  Indels   92;  Gaps   12;

Qy      7 NNNTSSSPGLFQSGDGNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNA 66
              :|:.||:| | |  :|                            ||:
Db     78 SNSTASSEGSVSSSNSG---------------------------SQSTSNSGSE 105

Qy     67 ATGA--GGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGG 124
```

```
Db    106 ASGSSNSGSQSTSNSGSEASGSSNSGSQSTDSSNSGSQSTGSSNSGSQSTSDSSNSGS 165
              |:|: |:|: |:|: :|:| |||:| :|:|: |:|:|:|:|: |:|:|: |:|: |
Qy    125 QIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNN-SASSGTSSSSG-SPPNDLSGGKAPS 182
                                                       :
Db    166 QSSTD-------------------SSNSGSQGSTGSSNSGSEASGSSNSGSESSGSSNSGSESSGSESSS 211
              :|:|:
Qy    183 GNSPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAGNS 242
              |:|:|  |  |:|   |:|:|:|  || | :|:| :|:|: |
Db    212 GSSNSGSESSGSSNSGSESSGSSNSGSESSGSSNSGSESSGSSNSGSESSGSSNS 271
              ||:|:
Qy    243 VAFTSAGANQTVLHDTITVKAGQVFDGKGQTFTAGSELGDGGQSENQKPLFI-LEDGASL 301
                  :|
Db    272 GSESSGSS----------NSGS--ESSGSSNSGSESSSDSGSSSDGKTTCISFHDTLSI 319
                  ||
Qy    302 KNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPN--SAGKKSHVEITNSSFEHAS 359
                  ||
Db    320 NTV----DD--------------DEIECTGKGETRCISDNNYKCATKQRHGSIECS------ 357
                   :|     .|
Qy    360 DKILQLNADTNLSVDNVKAK 379
                 .|      |
Db    358 -----VNGYIRCTGSNIKCK 372


RESULT 8
YK98_MYCTU
ID   YK98_MYCTU    STANDARD;    PRT;    434 AA.
AC   Q10707;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 36.5 KDA GLYCINE-RICH PROTEIN RV2098C.
GN   RV2098C OR MTCY49.38C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   -!- CAUTION: THIS PROTEIN IS FRAMESHIFTED NEAR THE THE N-TERMINUS.
CC       THE SEQUENCE HAS BEEN CHECKED BY AUTHORS IN REF.1 AND THEY REPORT
CC       THAT NO ERRORS HAVE BEEN FOUND.
CC   --------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; Z73966; CAA98228.1; -.
DR   HSSP; P19972; 1KVD.
DR   TubercuList; Rv2098c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR002952; -.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR01228; EGGSHELL.
```

```
KW   Hypothetical protein.
SQ   SEQUENCE   434 AA;  36530 MW;  B736E822CD1E0E54 CRC64;

Query Match
Best Local Similarity  25.3%;  Pred. No. 0.06;  Length 434;
Matches   80;  Conservative  17;  Mismatches 109;  Indels 110;  Gaps  13;

Qy   19 SGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAGGNDQTTG 78
          :||:|:|||:| |:                     |
Db  172 AGGNGGDGGHGGSGKAGG------------------SGGGSGFGQFGGAGGLLYG 208

Qy   79 VGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGGOIGDNPLLKAMLKL 138
          || || || ::  |  | |  ||| | |: :|||:| ||::|: ||:|
Db  209 NGGAAGSGGNGDAGTGVSSDGFAGLGGSGCRGGDAGLIGVGGGGGNGDPGLGARLFQ 268

Qy  139 IARMMDGQSDFGQPGTG----NNSASSGTSSSGGSPF----NDLSGGKAP-SGNSPSGN 189
          :  :    ::||::  |    ||:||||:||||||||        ::| :| ||:| ||
Db  269 V------GSRGGDGVGGWLYGDGGGDGGGNGGLPFIGSTNAGNGGSARLIGNGGAG- 320

Qy  190 YSPVSTFSPPSTPTSPTSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGI-GAGNSVAFTSA 248
                             |||              | | |  :| :||| | | |
Db  321 -------------------GSGGS----GAPGSVSSGGVGGAGNP---GGS 345

Qy  249 GANQTVLHDTITVKAGOVFDGKGQTFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGD 308
          |              |        :| | |  ||
Db  346 GGN--------GGVWYGNG--GAGGAAGQGGPGMN-----------TTSPGG 376

Qy  309 DGADGIH------LYGD 319
          ||   |:      |:||
Db  377 PGGVGHGGTAILFGD 392


RESULT  9
ID   YK83_MYCTU   STANDARD;   PRT;   314 AA.
AC   Q10691;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL 30.8 KDA PROTEIN RV2083.
GN   RV2083 OR MTCY49.22.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; Z73966; CAA98195.1; -.
DR   TubercuList; Rv2083.
```

```
KW   Hypothetical protein; Transmembrane.
FT   TRANSMEM      23     43       POTENTIAL.
FT   TRANSMEM      98    118       POTENTIAL.
FT   TRANSMEM     221    241       POTENTIAL.
SQ   SEQUENCE   314 AA;  30773 MW;  81E97B754EECED1E CRC64;

Query Match
Best Local Similarity  28.8%;  Pred. No. 0.05;  Length 314;
Matches   75;  Conservative  25;  Mismatches  71;  Indels  89;  Gaps  17;

Qy   16 LFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAA----TGA 70
          || |:  | ||| || :                ::    ||                |
Db   42 LFGLSTNADLGGANAD------IEGDDTDRRAHAADAARK--FSANEANAAEQMQGVGA 92

Qy   71 GGNDQ-TTGVGNA--GGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGGOIG 127
          ||      |:| :  ||:  :  | :   |                      |:| ||||:|
Db   93 QGMAQMASGIGGALSGALGGVMGPLTQLPQQAMQ-----------AGQGAMQ-- 133

Qy  128 DNPLLKAM--------LKLI-ARMMDGQSDFGQPGTGNNSASSGTSSSGGSPFNDLSG 177
          :|  : |         ::  |                                  ||
Db  134 --PLMSAMQQAQCADGLAAVDGARLLDSIG---GEPGLG--------SGAGG--GDVGG 177

Qy  178 GKAPSGNSPSGNYS--PVSTFSPPSTPTSPTSPLDFPSSPTKAA-----GGSTPVTDH-- 228
          | |  |    |      | :  | :  |  |   ||          :|: |
Db  178 GGA-GGTPTGYLGPPPVPTSSPPTPA---------GAPTKSATMPPPGGASPASAHMG 227

Qy  229 -------PDPVGSAGIGAG 240
             ::   | ||: |:|
Db  228 AAGMPMVPPGAMGARGEGSG 247


RESULT  10
ID   HYR1_CANAL   STANDARD;   PRT;   937 AA.
AC   P46591;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   HYPHALLY REGULATED PROTEIN PRECURSOR.
GN   HYR1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 10261;
RX   MEDLINE=96404781; PubMed=8808922;
RA   Bailey D.A., Feldmann P.J.F., Bovey M., Gow N.A.R., Brown A.J.P.;
RT   "The Candida albicans HYR1 gene, which is activated in response to
RT   hyphal development, belongs to a gene family encoding yeast cell wall
RT   proteins.";
RL   J. Bacteriol. 178:5353-5360(1996).
CC   -!- FUNCTION: NONESSENTIAL COMPONENT OF THE HYPHAL CELL WALL
CC       (POTENTIAL).
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC       (POTENTIAL).
CC   -!- DEVELOPMENTAL STAGE: ABUNDANT IN HYPHAE.
CC   -!- INDUCTION: INDUCED SPECIFICALLY IN RESPONSE TO HYPHAL DEVELOPMENT.
CC   --------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------
DR   EMBL; Z50123; CAA90485.1; -.
KW   Cell wall; Glycoprotein; Signal; GPI-anchor; Transmembrane; Repeat.
FT   SIGNAL        1      ?        POTENTIAL.
FT   CHAIN         ?    913        HYPHALLY REGULATED PROTEIN
```

```
FT   PROPEP        914    937    REMOVED IN MATURE FORM (POTENTIAL).
FT   TRANSMEM       42     62    POTENTIAL.
FT   DOMAIN        332    576    SER/THR-RICH.
FT   DOMAIN        569    576    POLY-THR.
FT   DOMAIN        577    814    ASN/GLY/SER-RICH.
FT   DOMAIN        610    753    7 X 4 AA REPEATS OF N-E-G-S.
FT   REPEAT        610    613    1.
FT   REPEAT        666    669    2.
FT   REPEAT        680    683    3.
FT   REPEAT        690    693    4.
FT   REPEAT        698    701    5.
FT   REPEAT        738    741    6.
FT   REPEAT        750    753    7.
FT   DOMAIN        901    904    POLY-SER.
FT   LIPID         913    913    GPI-ANCHOR (POTENTIAL).
FT   CARBOHYD       16     16    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      236    236    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      449    449    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      488    488    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      580    580    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      585    585    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      595    595    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      603    603    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      619    619    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      631    631    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      641    641    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      649    649    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      711    711    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      747    747    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      759    759    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      773    773    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      897    897    N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      913    913    N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE      937 AA;  93699 MW;  17802F121E5BA926 CRC64;

Query Match             6.9%;   Score 160.5;  DB 1;  Length 937;
Best Local Similarity  23.9%;   Pred. No. 0.17;
Matches  94;  Conservative  47;  Mismatches 135;  Indels 117;  Gaps  20;

Qy      5 TLNNNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSG  64
            :  ::||:       :|||| ||::|   ||: :  |||
Db    565 TVDATTTTTT---STGGDNSTGGNESGSNHG--PGNGST--------------EGSSG  602

Qy     65 NAATGAGGNDQT-----TGVGNAGGLNGRKGTAGTTPQSDSONMLSEMGNNGLDQAIT  117
          | ||| ||:   || :|||||| |
Db    603 N-GSGAGSNEGSQSGPGNNGSGSSEGGSNNGSGS----DSGSNNGSGSGSNNGSGSGST  656

Qy    118 PDGQGGQGIGDNPLLKAMLKLIARMMDGQSDQFGQP--------GTGNNSASS------GT  164
             ||| | :|
Db    657 EGSEGGS--GSNE------GSQSGSGSQPGPNEGSEGGSGSNEGSNHGSNEGS  701

Qy    165 SSSGGSPFNDLSGGKAPSGN--SPSGNYSPVSTFSPPSTPTSPLDFPSSPTKAAGG  221
          || | | ::|
Db    702 GSGSGSNNGSGSGSQSGSGSGSQSGSES---------------GSNSGSNEG  740

Qy    222 STPVTDHPDPVGSAGIGAGNSVAFTSAGANQTVLHDTITVKAGVFDGKGQTFTAGSELG  281
          : | :|:
Db    741 SNPGAGNGSNEGS-GQGSGNG---SEAGSGQGSGNNGSGSGHN-DGSGSGSNQGSNPG  794

Qy    282 DGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNVGEDAITVKPN-  340
          ||  :      :||  ::|
Db    795 AGSGSGS------ESGSKAGSHSGSNEGAK-----TDSIEGFHTES  833

Qy    341 SAGKKSHVEITNSSFEHASDKILQLNADTNLSV  373
          : |:  : |::     |:  ||:|||::||||
Db    834 NTGAHTDATVTGNSV--ANPVTTSTESDTTISV  864

RESULT 11
Y278_MYCTU
ID   Y278_MYCTU    STANDARD;    PRT;    957 AA.
AC   P56877;
```

```
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV0278C PRECURSOR.
GN   RV0278C OR MTV035.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares R., Squares S., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021930; CAA17353.1; -.
DR   TubercuList; Rv0278c; -.
DR   InterPro; IPR000084; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal.
FT   SIGNAL         1     30    POTENTIAL.
FT   CHAIN         31    957    HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                             RV0278C.
SQ   SEQUENCE      957 AA;  81905 MW;  71EBABD417FBA47C CRC64;

Query Match             6.9%;   Score 160.5;  DB 1;  Length 957;
Best Local Similarity  25.5%;   Pred. No. 0.17;
Matches  92;  Conservative  31;  Mismatches 117;  Indels 121;  Gaps  17;

Qy      8 NNTSSSPG--------LFQSGGDNGLGGHN-ANSALGQQPIDRQTIEQMAQLLAELLK  56
          | : :||         ::|| | || :
Db    555 NGANGTPGTGAAGGAGGWLFGNGNGGHGAINTAATATGG--------AGGAG  599

Qy     57 SLLSPQSGNA----ATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSEMGNNG  111
          ::|| | |     | |||| :  :|| |||| |||           |:    ::|||
Db    600 GILFGTGNGGTGGIATGAGG---IGGAGGAGGVSLLIGSGGTG-----------GNGG  644

Qy    112 LDQAITPDGQGGQGIGDNPLLKAMLKLIARMMDGQSDQFGQPTGNNSAS---SGTSSS  167
          :   || |  |||| | :      ||    |  |:   |:    |::     |:|:||
Db    645 NSIGVAGIGGAGGRGGDAGLL----------FGAAGTGGHGAAGGVPAGVGGA----SGTSSS  687

Qy    168 GGSPFNDLSGGKAPSG--NSPSGNYSPVSTFSPPSTPTSPTS---------PLDFPSSP  215
          || |  :  | || |   |  :| |  :: : | | | | |         || ||:
Db    688 GGNGGLFANGGAGGAGGFNAAGGNGGNGGLFGTGGGAGGNGGNGGL-----FGAGG  746

Qy    216 TKAAGGSTPVTDHPDPVGSAGI-----GAGNSVAFTSAGANQTVLHDTITVKAGQVFDGK  270
          | ||||       |  ||| :      |||| |  :
Db    747 TGGAAGS-------GGSGITTGGGHGGNAGLLSLGA  784

Qy    271 GQTFTAGSELGDGGQSENQKPLFIL---EDGASLKNVTMGDDGADGIH---LYGD  319
          |  : ||  |||| :      ||       |  ||   ||| || |:   ||||
Db    785 GASSLAG--GAGGTGGNGALLFGFRGAGGAGGHGGAALTSIQQGGAGGAGGNGGLLFGS  841
```

```
Qy   320 A 320
             |
Db   842 A 842
```

```
RESULT 12
YQ34_MYCTU
ID   YQ34_MYCTU     STANDARD;      PRT;    778 AA.
AC   P71933;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2634C.
GN   RV2634C OR MTCY441.04C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
CC       SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z80225; CAB02341.1; -.
DR   HSSP; P04002; 1ATF.
DR   TubercuList; Rv2634c; -.
DR   InterPro; IPR000084; -.
DR   InterPro; IPR001899; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   778 AA;  63131 MW;  DAB20FE58E4999E7 CRC64;

Query Match             6.9%;  Score 159.5;  DB 1;  Length 778;
Best Local Similarity  24.2%;  Pred. No. 0.15;
Matches   82;  Conservative  26;  Mismatches 128;  Indels 103;  Gaps 16;

Qy    15 GLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGNAATGAG--G 72
             |:|| ||  | | |||  | :      |||     |||||||
Db   252 GVFGNGGFGGAGGLGAAGGVGG----------------AASYFGTGGGGVGGDGAPG 293

Qy    73 NDQTTG---VGN--AGGLNGRKGTAGTTPQSDSQNMLSEMGNNGLDQAITPDGQGGGQIG 127
            ||      |||    ||    ||      ||||  | |||     ||  |||
Db   294 GDGGAGPLLIGNGGVGGLGG-AGAAG------------GNGGAGGMLLGDGGAGGQGG 338

Qy   128 DNPLLKAMLKLIARMMDGQSDQFGQPGTGNNSASSGTSSSGGSPFNDLSGGKAPSGNSPS 187
            |  :  :       :|||       |       :|| :::  ||| |  |  |||
Db   339 --PAVAGVLG----------GMPGAGGNGGNANWFGSGG------AGGQGGTGLAGT 377

Qy   188 GNYSPVSTFSPPSTPTPSTPSPLDFPSSPTKAAGGSTPVTDHPDPVGSAGIGAGNSVAFTS 247
            | :    : :     :|   ||| :| |
```

```
Db   378 NGVNP-GSIANPNTGANGTD--------NSGNGNQTGGNGGPGPAG--GVGEAGGVGGQG 426
                          :   |||            |||     | | |
Qy   248 AGANQTVLHDTITVKAGQVFDGK------GQFTAGSELGDGGQSENQKPLFILEDGAS 300
                          ||                  |||::|||  |
Db   427 G------------LGESLDGNDGTGGKGAGGTAGTDGGAGGAGGAGG---IGETDGS 469

Qy   301 LKNVTMGDDGADGIHLYGDAKIDNL------HVTNVGE 332
              |  :||| :|| |:|           | |:|
Db   470 AGGVATGGEGGDGATGGVGGVGGAGGKGGQGHNTGVGD 508
```

```
RESULT 13
YS89_CAEEL
ID   YS89_CAEEL     STANDARD;      PRT;    2329 AA.
AC   Q09624;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HYPOTHETICAL 254.3 KDA PROTEIN ZK945.9 IN CHROMOSOME II.
GN   ZK945.9.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z48544; CAA88442.1; -.
DR   WormPep; ZK945.9; CE01740.
DR   InterPro; IPR000203; -.
DR   InterPro; IPR001024; -.
DR   Pfam; PF01825; GPS; 1.
DR   Pfam; PF01477; PLAT; 1.
KW   Hypothetical protein; Repeat; Transmembrane.
FT   DOMAIN        22    115       SER/THR-RICH.
FT   DOMAIN       250    389       GLY/SER-RICH.
FT   TRANSMEM     557    577       POTENTIAL.
FT   TRANSMEM     606    626       POTENTIAL.
FT   TRANSMEM    1162   1182       POTENTIAL.
FT   TRANSMEM    1290   1310       POTENTIAL.
FT   TRANSMEM    1467   1487       POTENTIAL.
FT   TRANSMEM    1498   1518       POTENTIAL.
FT   TRANSMEM    1541   1561       POTENTIAL.
FT   TRANSMEM    1602   1622       POTENTIAL.
FT   TRANSMEM    1637   1657       POTENTIAL.
FT   TRANSMEM    1717   1737       POTENTIAL.
FT   TRANSMEM    1935   1955       POTENTIAL.
FT   TRANSMEM    1990   2010       POTENTIAL.
FT   TRANSMEM    2039   2059       POTENTIAL.
FT   TRANSMEM    2088   2108       POTENTIAL.
FT   TRANSMEM    2134   2154       POTENTIAL.
FT   TRANSMEM    2189   2209       POTENTIAL.
SQ   SEQUENCE   2329 AA;  254348 MW;  4D9D423060F0CA85 CRC64;

Query Match             6.8%;  Score 157.5;  DB 1;  Length 2329;
Best Local Similarity  20.8%;  Pred. No. 0.65;
Matches  103;  Conservative  62;  Mismatches 201;  Indels 129;  Gaps 17;

Qy     3 ILTLN---NNTSSSPGLFQSGGDNGLGGHNANSALGQQPIDRQTIEQMAQLLAELLKSL 58
            | |:|           |             ||  | |  |  |  |
```

```
Db  230 IFTMNVVLPTTTTETPPTTVSSSDDAGGKTGGTGATG----------------------- 266

Qy   59 LSPQSGNAATGAGGNDQTTGVGNA------GGLNGRKGT------AGTTPQSDSQNMLSE 106
            ||:||::  :|: ||          :||          .:
Db  267 ----GTGGTGSGSATTLSTGDAVRSTTSGSGSGQSSTGSGAGGSGTTASGSGSGGSSG 321

Qy  107 MGNNGLDQ----AITPDGQG------GGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGT 155
            ::||:::      :: ||        :|
Db  322 TGSDGVNSGKTTALNGDGTGSGTATTPGSHLGD------------------------ 354

Qy  156 GNNSASSGTSSSGGSPFN-DLSGGKAPSGNS-PSGNYSPVSTFSPPSTPTSPTSPLDFPS 213
            | |:||  ||:  ||||| ::|: |:||||:  |||  |::     ||:|| ||||
Db  355 GGSTSGGSGDSNGSSGGVSTKSSSGSGSDTSGSSDSSGANGAFSATAQPSTRTTKTRSSLATV 414

Qy  214 SPTKAAGGSTPVTDHPDPVGS-AGIGAGNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQ 272
           | |:::     |   ||:|| : |||  :||   |:::     | ||::||::||::
Db  415 SPISAARQAIIDAQKADVMNQLAGIMDG-SASNNSLNTSSSLLNQISSLPAADLVEVAQS 473

Qy  273 TFTAGSELGDGGQSENQKPLFILEDGASLKNVTMGDDGADGIHLYGDAKIDNLHVTNV-- 330
           |||||| |:|: ||||| :|:|: : | |:::|| || ||    | | :|   ::|
Db  474 LLSNTLKIPGVGNMSSVDVLKTLQDNIATTNSELADEMAKVI----TKLANVNMTSAQS 528

Qy  331 --------GEDAITVKPNSAGKKSHVEITNSSFEHAS------------DKIL 363
                    :||  : |:  ||  :|   |  :
Db  529 LNSVLSSLDLALKGSTVYTLGVSSTKSKDGTYAVIFGVVIASCYTLVSPRCTLSIYGSTI 588

Qy  364 QLNADTNLSV-----DNVKAKDFGTFVRTNGGQOGNWDLNLSHISAEDGKFSEFVKSDSEG 418
            |: :  :|       |: ||| : :: | |: |||::    ||:|   :|
Db  589 YLTGDTRASYKQLDGDTVTA---DTMLAAAIGIQGMFATNGRTVQVEQDKIDDKRSLVSG 645

Qy  419 LNVNTSDISLGDVEN 433
           :|:||:  :|||
Db  646 -NIMATMSGVGDVQS 659


RESULT 14
YP91_MYCTU
ID   YP91_MYCTU    STANDARD;    PRT;   543 AA.
AC   Q50630;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HYPOTHETICAL PE-PGRS FAMILY PROTEIN RV2591.
GN   RV2591 OR MTCY227.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA   Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
-!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY; PGRS
    SUBFAMILY.
-----------------------------------------------------------------------
```

```
CC  -----------------------------------------------------------------------
DR   EMBL; Z77724; CAB01283.1; -.
DR   TubercuList; Rv2591; -.
DR   InterPro; IPR000084; -.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   543 AA;   46287 MW;  597303395D2DF59 CRC64;

Query Match             6.8%;  Score 157;  DB 1;  Length 543;
Best Local Similarity   23.1%;  Pred. No. 0.14;
Matches   81;  Conservative  39;  Mismatches 143;  Indels 88;  Gaps 14;

Qy    1 MSILTLNNNTSSSSPGLFQSGGDN-----------GLGGHNANSALGQOPIDRQTIEQMAQLLAEL 54
          ||:|       :|:   :|::|                  :  ||: :
Db    1 MSFVT------AAPEMLATAAQNVANIGTSLSAANATAAASTTSVLAAGADEVSQAIARL 54

Qy   55 LKSLLSP-QSGNAATGAGGNDQTTGVGNAGGLNGRKGTAGTTPQSDSQNMLSE--------- 106
          |:|:||   ||  :|| :|:      || :| |:
Db   55 FSDYATHYQSLNAQAAAFHHSFVQTLNAAGGAYSSAEAANASAQALEQNLLAVINAPAQA 114

Qy  107 -------MGNNGLDQAITPDGQGGQGI--GDNPLLKAMLKLLIARMMDGQSDQFGQPGTGN 157
                 :|| |   || |:||:  |:||  :  | |: |
Db  115 LFGRPLIGNGANGTAASPNGGDGGILYGNGGN----GFSQTTAGVAGGSAGGSAGLIGNGG 170

Qy  158 NSASSGTSSSGGSPPNDLSGGKAP--SGNSPSGNYSPVSTFSPPSTPTSPLDFPSSP 215
          || ||| ||||::   :||| |    |||:||  |||: || ||:|| ||||:
Db  171 NGGAGGAGAGG------AGGAGGWLLGNGGAG------------------GPGGPTDVPAGT 209

Qy  216 TKA-AGGSTPVTDHPDPVGSAGIGA-GNSVAFTSAGANQTVLHDTITVKAGQVFDGKGQ 272
          |:|  ||:|                  | : |:  ||
Db  210 GGAGGAGGDAPL----------IGWGGNGGPGGFAAFGNGG------------AGGNGGASGS 250

Qy  273 TFTAGSELGDGGQSEN-----------QKPLFILEDGASLKNVTMGDDGADG 313
          |||| :||||||:                 :| | :|||||||||||
Db  251 LFGVGGAGGVGGSSEDVGGVGGTGGAGGRGGLFLGLGDGGAGGTSNNGGDG 301


RESULT 15
N116_YEAST
ID   N116_YEAST    STANDARD;    PRT;   1113 AA.
AC   Q02630;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   NUCLEOPORIN NUP116/NSP116 (NUCLEAR PORE PROTEIN NUP116/NSP116).
GN   NUP116 OR NSP116 OR YMR047C OR YM9532.12C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93054906; PubMed=1385442;
RA   Wente S.R., Rout M.P., Blobel G.;
RT   "A new family of yeast nuclear pore complex proteins.";
RL   J. Cell Biol. 119:705-723(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93099880; PubMed=1464327;
RA   Wimmer C., Doye V., Grandi P., Nehrbass U., Hurt E.C.;
RT   "A new subclass of nucleoporins that functionally interact with
RT   nuclear pore protein NSP1.";
RL   EMBO J. 11:5051-5061(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RA   Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: ESSENTIAL COMPONENT OF NUCLEAR PORE COMPLEX.
CC       NUCLEOPORINS MAY BE INVOLVED IN BOTH BINDING AND TRANSLOCATION OF
CC       THE PROTEINS DURING NUCLEOCYTOPLASMIC TRANSPORT.
CC   -!- SUBUNIT: INTERACTS WITH KAP95.
```

Job time: 127 sec

```
CC   -!- SUBCELLULAR LOCATION: NUCLEAR PORE COMPLEX.
CC   -!- DOMAIN: CONTAINS G-L-F-G REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE GLFG FAMILY OF NUCLEOPORINS.
CC   ------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------
DR   EMBL; Z15036; CAA78754.1; -.
DR   EMBL; X68108; CAA48228.1; -.
DR   EMBL; Z48502; CAA88413.1; -.
DR   PIR; S28537; S28537.
DR   PIR; S28925; S28925.
DR   PIR; A44402; A44402.
DR   SGD; S0004650; NUP116.
KW   Nuclear protein; Transport; Repeat.
FT   DOMAIN        205    715       37 X 6 AA APPROXIMATE REPEATS OF
FT                                  G-L-F-G.
FT   CONFLICT       26     26       G -> A (IN REF. 1).
FT   CONFLICT      536    536       S -> G (IN REF. 1).
FT   CONFLICT      720    720       S -> P (IN REF. 1).
FT   CONFLICT     1018   1018       S -> Y (IN REF. 1).
FT   CONFLICT     1023   1023       I -> Y (IN REF. 1).
SQ   SEQUENCE   1113 AA;  116234 MW;  FBAB0B9AEA958213 CRC64;
```

```
Query Match             6.7%;  Score 155;  DB 1;  Length 1113;
Best Local Similarity  23.0%;  Pred. No. 0.39;
Matches  115;  Conservative  60;  Mismatches 180;  Indels  144;  Gaps  26;

Qy      11 SSSPGLF----QSGGDNGL-GGHNANSALGQQPIDRQTIEQMAQLLAELLKSLLSPQSGN 65
           |:::  ||    |    :| |          :|         ::  ::    | |::|:
Db     403 SNAGGLFGQNNQSQNQSGLFGQQNSSNAFGQP--------QQQGGLFGSKPAGGLFGQQQG 455

Qy      66 AATGAGGNDQTTGV-------GNAGGLNGRK--------GTAGTTPQDSQNMLSEMG 108
           |:|/|||| |         |||:|  :||:|        |:::|
Db     456 ASTFASGNAQNNSIFGQNNQQQSTGGLFGQQNNQSQSPGGLFGQTNQNNNQ----PFG 511

Qy     109 NNGLDQAITPDGQGGQIGDNPLLKAMLKLIARMMDGQSDQFGQPGTGNN---------- 158
            |||    |  |        |  |   :       |  :||   :|||
Db     512 QNGLQQ----PQQNNSLFGAKPTGFGNTSLFSNSTTNQSNGI----SGNNLQQQSGLFQ 563

Qy     159 ----SASSGTSSSGGSPFNDLSGGKAPSGNSPSGNY-SPVSTF----SPPST-----PT 203
               :    :  |   |   |    ||  |   |  :  |  ||    | |:|     | |
Db     564 NKQQPASGGL--FGSKPSNTVGGGL--FGNNQVANQNNPASTSGGLFGSKPATGSLFGGT 619

Qy     204 SPTSPLDFPSSPTKAAGG-------STPVTDHPDPVGSAGIGAGNSVAFTSAGA---- 250
           | :| |       ||||         | :|:    :| | |  | |     | |
Db     620 NST------APNASSGGIFGSNNASNTAATTNSTGLFGNKPVGAG---ASTSAGGLFGN 669

Qy     251 -NQTVLHD---------TITVKAGQVFDGKGQTFTAGSEL------GDGGQSENQKP 291
            :|:              |:| |  | |  | ::|  |          |:| ::
Db     670 NNNSSLNNSNGSTGLFGSNNTSQSTNAGGLFQNNTSTNTSGGGLFSQPSQSMAQSQN--- 726

Qy     292 LFILEDGASLKNVTMGDDGADGIH-LYGDAKIDNLHVTNVGEDAITVKPNSAGKKS---- 346
           |    | | :   | ||| |      |:|||  :|  |   |  |   |    ||:
Db     727 --ALQQQQQQRLQIQNNNPYGTNELFSKATVTNTVSYPIQPSATKIKADERKKASLTNA 784

Qy     347 -------HVEITNSSFEHASDKILQLNADTNLSVDNVKAKDFGTFVRTNGGQGGNW 395
                       |: ||   | :    |  :  :| |  |       |:| : :|:|
Db     785 YKMIPKTLFTAKLKTNNSVMDKA-----QIKVDPKLSISIDKKNN---QIAISNQQEENL 836

Qy     396 D---LNLSHISAEDGKFSF 411
               |::|:|| :|| ||:|
Db     837 DESILKASELLFNPDKRSF 855
```

Search completed: October 18, 2001, 09:10:53